**Published quarterly by the Institute of Mathematical Statistics in March, June, September and December.**


## IMS INSTITUTIONAL MEMBERS

ABERDEEN PROVING GROUNDS (BALLISTIC RESEARCH LABORATORIES), Aberdeen, Maryland
AMERICAN VISCOSE CORPORATION, Marcus Hook, Pennsylvania
BELL TELEPHONE LABORATORIES, INC., TECHNICAL LIBRARY, 463 West Street, New York 14, New York
CALIFORNIA RESEARCH CORPORATION, P. O. Box 1627, Richmond, California
CORNELL UNIVERSITY, MATHEMATICS DEPARTMENT, Ithaca, New York
GENERAL ANALYSIS CORPORATION, 11753 Wilshire Blvd., Los Angeles 25, California
INDIANA UNIVERSITY, THE LIBRARY, Bloomington, Indiana
INTERNATIONAL BUSINESS MACHINES CORPORATION, APPLIED SCIENCE LIBRARY, White Plains, New York
IOWA STATE UNIVERSITY, STATISTICAL LABORATORY, Ames, Iowa
LOCKHEED AIRCRAFT CORPORATION, ENGINEERING LIBRARY, Burbank, California
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, HAYDEN LIBRARY, PERIODICAL DEPARTMENT, Cambridge 39, Massachusetts
MICHIGAN STATE UNIVERSITY, DEPARTMENT OF STATISTICS, East Lansing, Michigan
NATIONAL SECURITY AGENCY, Fort George G. Meade, Maryland
PRINCETON UNIVERSITY, DEPARTMENT OF MATHEMATICS, SECTION OF MATHEMATICAL STATISTICS, Princeton, New Jersey
PURDUE UNIVERSITY LIBRARIES, Lafayette, Indiana
RADIO CORPORATION OF AMERICA, R.C.A. LABORATORIES, Princeton, New Jersey
SANDIA CORPORATION, Sandia Base, Albuquerque, New Mexico
SPACE TECHNOLOGY LABORATORIES, P. O. Box 95001, Los Angeles 45, California
STANFORD UNIVERSITY, GIRSHICK MEMORIAL LIBRARY, Stanford, California
STATE UNIVERSITY OF IOWA, Iowa City, Iowa
THE CATHOLIC UNIVERSITY OF AMERICA, ENG. AND MATH. LIBRARY, Washington, D. C.
THE RAMO-WOOLRIDGE CORPORATION, Los Angeles, California
UNION CARBIDE CORPORATION, 30 East 42nd Street, New York 17, New York
UNITED STATES STEEL CORPORATION LIBRARY, Monroeville, Penna.
UNIVERSITY OF CALIFORNIA, STATISTICAL LABORATORY, Berkeley, California
UNIVERSITY OF ILLINOIS, SERIALS DEPARTMENT, Urbana, Illinois
UNIVERSITY OF NORTH CAROLINA, DEPARTMENT OF STATISTICS, Chapel Hill, North Carolina
UNIVERSITY OF PUERTO RICO, SCHOOL OF TROPICAL MEDICINE, San Juan, Puerto Rico
UNIVERSITY OF WASHINGTON, LABORATORY OF STATISTICAL RESEARCH, Seattle, Washington

[illegible]

# SOME VALIDITY CRITERIA FOR STATISTICAL INFERENCES

By Robert J. Buehler

*Iowa State University*

**1. Introduction.** This paper is concerned with the ways in which existing statistical theories specify the degree of uncertainty of an inference. For the sake of graphic presentation, problems of inference are described in terms of a game between two players—one who makes the inferences and another who questions their validity. Such a model suggests a number of criteria of validity which depend entirely on classical probability calculations. I anticipate that arguments may be advanced for *not* regarding statistical inference as a game; it is hoped that the cogency of such arguments will not prevent the present model from providing some new insight into the problems here considered.

Many, though not all, problems of inference lead to assertions of the type, "The probability that $A$ is true is equal to $\alpha$," or, "$P(A) = \alpha$." One may ask whether the person making this assertion should be willing to bet that $A$ is true, risking an amount $\alpha$ to win $1 - \alpha$, and should be equally willing to bet that $A$ is false, risking $1 - \alpha$ to win $\alpha$, against an opponent who has exactly the same information as he and who is allowed to choose either side of the wager. The affirmative answer will not be defended here, but its consequences will be examined.

The game viewpoint is related to, but not identical with, the ideas of von Mises, who has advanced as a postulate "the impossibility of a gambling system" in his definition of probability (see for example [19], p. 15). It has generally been recognized that modern theories of inference, which avoid the assumption of prior distributions of the parameters, should not have the same interpretation as the classical Bayes-Laplace theory based on prior distributions. The present paper attempts to show the sense in which one pays for weakening the classical assumptions by losing the von Mises postulate for the inferences "$P(A) = \alpha$."

Sections 2 to 5 are devoted to the theory of confidence intervals; in Sections 6 to 9 the ideas are generalized to include other statistical problems. The reader is warned not to expect to find any new problems solved in this paper, for at the present stage of development the theory gives at best new ways of looking at existing solutions.

**2. A model for studying interval estimation.** In the spirit of the introductory section the problem of interval estimation is here studied in terms of a game between two players. The players have equal knowledge about fixed conditions $K$ (for "known") of a random experiment, for example, knowledge that $n$ values are observed from a normal population having unit variance. Unknown conditions of the experiment, for example the value of the population mean $\mu$, are conveniently referred to as the "state of nature" $U$ (for "unknown"). The

Received October 21, 1958; revised June 19, 1959.

first player, Peter, has the familiar task of setting confidence intervals. It is required that he formulate a rule $R$ which determines the interval as a function of the observations. Then on the basis of the observations he makes a *probability assertion*

$$\text{“}P(A) = \alpha\text{”}.$$

The quotation marks are used to identify an expression as Peter's assertion and to warn that it may not have validity in a direct probability or frequency sense, for that restriction is not imposed. The assertions may have validity as "confidence probabilities" or "fiducial probabilities," these being special cases. In the unit-normal example one would commonly take $\alpha$ to be 0.95 and $A$ to be

$$\bar{x} - 1.96/\sqrt{n} \leqq \mu \leqq \bar{x} + 1.96/\sqrt{n}.$$

In order that the second player, Paul, have information equal to Peter's, it is required that he have knowledge of Peter's rule $R$ as well as of the experimental conditions and observations. Paul adopts a strategy $S$ based on $R$ and on the experimental conditions, and consisting in the specification of two subsets $C^+$ and $C^-$ of the observation space such that

$$\text{for observations in} \begin{cases} C^+ \\ C^- \end{cases} \text{Paul bets that}$$

$$A \text{ is} \begin{cases} \text{true,} \\ \text{false,} \end{cases} \quad \text{risking} \begin{cases} \alpha \\ 1 - \alpha \end{cases} \quad \text{to win} \begin{cases} 1 - \alpha. \\ \alpha. \end{cases}$$

It is not required that a bet must always be made; thus $C^+$ and $C^-$ need not be exhaustive. To determine the winner of each bet, we postulate the existence of a referee who knows the true state of nature.

2.1. *The criterion of weak exactness.* If the model is adequately specified, one should in principle be able to calculate the expected gain to Paul. For any fixed experimental conditions $K$ the expected gain would be a function of (i) the state of nature $U$, (ii) Peter's rule $R$, and (iii) Paul's strategy $S$. Different criteria for the sensibility of Peter's rule might be put forward in terms of this expected gain. For example, I propose the following. Suppose Paul's strategy is to bet consistently that $A$ is false, regardless of the observations. Then if Paul's expected gain is zero for all $U$, Peter's rule $R$ will be defined to be *weakly exact.*[1]

Weak exactness is (at least from the point of view of some theories of probability) essentially equivalent to writing $P(A) = \alpha$ without quotation marks, the definition above being preferred on the grounds that it is less subject to misinterpretation. It will be noted that weak exactness is a propery possessed by Neyman's confidence intervals (see for example [17]), but not necessarily by the fiducial counterparts, as is known.

2.2. *Relevant and semirelevant subsets.* In this paper the ultimate calculation of the expected gain is actually made only once (in Section 4.2). The emphasis lies

[1] Strongly exact is defined in Section 6.1.

more in studying Paul's initial search for a strategy having a guaranteed winning percentage. If we call $P(A \mid C) - \alpha$ the *bias* of $C$, then Paul's problem is to find subsets $C$ whose bias has the same sign for all $U$. These will be called *semirelevant subsets induced by* $R$. If moreover the bias is bounded away from zero, they will be called *relevant*. That is, if $\epsilon > 0$ is independent of $U$, then $C$ is called

$$\text{semirelevant}\begin{cases} \text{if} & P(A \mid C) > \alpha \\ \text{or if} & P(A \mid C) < \alpha \end{cases} \quad \text{for all } U,$$

$$\text{relevant}\begin{cases} \text{if} & P(A \mid C) \geqq \alpha + \epsilon \\ \text{or if} & P(A \mid C) \leqq \alpha - \epsilon \end{cases} \quad \text{for all } U.$$

The phrase "induced by $R$" is crucial; the defined properties are necessarily relative to the rule $R$, which enters the defining equations through $A$. The definitions are inspired largely by writings of Fisher, of which the following quotations from [11] are typical:

p. 32. "···no such subset can be recognized."

p. 33. "···inability to discriminate any of the different subaggregates having different limiting frequency ratios."

p. 57. "···every subset to which it belongs, and which is characterized by a different fraction must be unrecognizable."

More recently ([12], p. 23) Fisher uses the words "relevant" and "irrelevant." For example:

"The subset of throws made on a Tuesday is easily recognizable, it has, however the same probability as the whole set and is therefore irrelevant."

It would seem that any subset $C$ of the observation space might be called "recognizable" in Fisher's sense since it is determined by known observations. The word "relevant" has been introduced here in a sense intended to be close to Fisher's, but there seems to be at least a small difference arising from the dependence on the rule $R$. If a need for distinct terms should arise, "induced relevant subset" might be substituted for "relevant subset" as used here.

In typical interval estimation problems the bias of most subsets $C$ will not have the same sign for all $U$. In particular if $C$ contains only a single point of the observation space, then $P(A \mid C)$ ordinarily will be either zero or unity, depending on $U$. This corresponds to the statement, sometimes seen in textbooks, that the probability that the true value of the parameter lies within a confidence interval is either zero or unity *after* the interval has been constructed.

2.3. *Relevance of unions and complements.* It will be convenient later to refer to the following elementary results.

LEMMA 1. *If subsets $C_1$ and $C_2$ are disjoint, (semi)relevant, and positively [negatively] biased, and if the union $C_1 + C_2$ has nonzero probability for all $U$, then the union is (semi)relevant and positively [negatively] biased.*

We may remark that for nondisjoint subsets neither the union nor the intersection need have special properties of relevance deriving from the components.

LEMMA 2. *Let $C'$ denote the complement of $C$. If* (i) $P(A) = \alpha$ *(essentially*

*weak exactness*); (ii) *$P(C)$ and $P(C')$ are nonzero for all $U$*; (iii) $P(C) > b > 0$ (*$b$ is a constant independent of $U$*); *and* (iv) *$C$ is relevant; then $C'$ is relevant with bias opposite to $C$.*

LEMMA 3. *If in Lemma* 2 (iii) *is not assumed, then $C'$ is semirelevant with bias opposite to $C$.*

LEMMA 4. *If in Lemma* 2 (iii) *is not assumed and if* (iv) *is replaced by "$C$ is semirelevant," then $C'$ is semirelevant with bias opposite to $C$.*

**3. Examples of relevant subsets in interval estimation problems.** We now give an assortment of six examples of relevant subsets induced by systems of confidence intervals.

3.1. *An example of intervals based on an insufficient statistic.* This first example, although relatively simple, illustrates a number of interesting points. Let a sample of size $n = 2$ be drawn from a normal population having unknown mean $\mu$ and unit variance. A confidence interval based on the first observation $x_1$, but ignoring the second $x_2$, corresponds to the weakly exact probability assertion

$$\text{“}P(x_1 - 1.96 \leqq \mu \leqq x_1 + 1.96) = 0.95\text{”}.$$

Paul might logically begin his search for relevant subsets by comparing this questionable assertion with the standard one based on the sufficient statistic $\bar{x} = \frac{1}{2}(x_1 + x_2)$. The comparison shows that when $x_1 = x_2$ the intervals are correctly placed but unduly long, and when $|x_1 - x_2|$ is large they are "badly placed." A clue is thus furnished which suggests conditioning on subsets defined in terms of the statistic $\delta = x_1 - x_2$. If $C$ denotes any such subset and $A$ denotes $|x_1 - \mu| \leqq 1.96$, then the conditional probability is

$$P(A \mid C) = P(AC)/P(C) = \iint_{AC} f\,dx\,dy \Big/ \iint_C f\,dx\,dy$$

where $f\,dx\,dy$ is the density $(1/2\pi)\exp\{-\frac{1}{2}(x_1 - \mu)^2 - \frac{1}{2}(x_2 - \mu)^2\}\,dx\,dy$. On setting $y_1 = x_1 - \mu$, $y_2 = x_2 - \mu$, it is seen that both numerator and denominator are independent of $\mu$. Thus $P(A \mid C)$ is independent of $\mu$, and all subsets $C$ defined in terms of $\delta$ will be relevant save for exceptional cases for which $P(A \mid C) = 0.95$. In particular the sets (of zero probability) for which $\delta = \delta_0$ can be seen by simple calculation to have negative, zero, positive bias respectively for $|\delta_0| >, =, < 1.593$. For $\delta_0 = 0$, $P(A \mid C)$ achieves its maximum value of 0.997 ($=$ standard normal area within $\pm 1.96\sqrt{2}$); as $|\delta_0|$ tends to infinity, $P(A \mid C)$ tends to zero. Consider now the three subsets

$$C_1\colon\ 0 < \delta < 1; \qquad C_2\colon\ -1 < \delta < 0; \qquad C_2'\colon\ \text{complement of } C_2 .$$

$C_1$ may be regarded as being made up of a continuum of positively biased relevant subsets of zero probability, and by a continuous generalization of Lemma 1 it follows that $C_1$ has positive bias. By the same reasoning $C_2$ is positively biased; and by Lemma 2, $C_2'$ is negatively biased. One interesting observation is that $C_1$ is a subset of $C_2'$; thus a positively biased subset may sometimes be contained in a negatively biased one. Furthermore it is possible for a particular observation

to belong simultaneously to two different subsets having opposite bias. Thus arises the basic question (which will not be resolved in this paper) of the *appropriate* subset to which any particular observation should be referred if it is not to be referred to the universal set of all observations.

3.2. *An example involving shortest average length.* Our second example is inspired by Cox [4], who treats the *testing* situation and the criterion of maximum power, whereas we treat the *estimation* situation and the criterion of minimum expected length. Suppose two populations are known to have standard deviations $\sigma_1 = 1$ and $\sigma_2 = 2$. A random choice between the two populations is made and one normal random variate is observed. If Peter knows which population was sampled he may treat each population separately in the usual way, asserting

$$\text{“}P(x - 1.96\sigma_i \leqq \mu \leqq x + 1.96\sigma_i) = 0.95\text{”}$$

in which both $x$ and $\sigma_i$ represent observed values. On the other hand, if he wished to minimize the average length of the intervals, he would do best to increase the error rate for the second population to about seven per cent and compensate by decreasing the error rate for the first population to about three per cent, thus maintaining a five per cent average. The first solution may be called a "conditional" one (in the sense of conditional probability) inasmuch as $P(A) = \alpha$ is valid in the frequency sense, with the relative frequency $\alpha$ prevailing separately in the two subsets defined by the value of $\sigma$. The second solution is "unconditional" in that $\mathrm{P}(A) = \alpha$ is valid in the frequency sense only when related to the sequence of all observations, ignoring the observed value of $\sigma$. Both solutions are weakly exact. The two subsets defined by the value of $\sigma$ are conspicuously relevant for the unconditional solution; thus the criteria of shortest average length and nonexistence of relevant subsets cannot both be met. I myself would prefer the conditional solution here.

3.3. *An example of relevant subsets defined by an ancillary statistic.* The following example uses some distributional results given by Fisher ([11], pp. 163–165), to illustrate how relevant subsets may be defined by means of an *ancillary* statistic. Consider a sample of size $n$ from the bivariate density

$$e^{-\theta x - y/\theta}\, dx\, dy, \qquad 0 < x, y, \theta < \infty,$$

and define statistics $T$ (the maximum likelihood estimator of $\theta$) and $V$ (an ancillary statistic, according to Fisher) by

$$X = \sum x, \qquad Y = \sum y, \qquad T^2 = Y/X, \qquad V^2 = XY.$$

Then $T$ and $V$ have the joint density

$$\frac{V^{2n-1}}{[(n-1)!]^2} \exp\left\{-V\left(\frac{T}{\theta} + \frac{\theta}{T}\right)\right\} \frac{dT}{T}\, dV,$$

and the marginal and conditional densities of $T$ are respectively

$$K\left(\frac{T}{\theta} + \frac{\theta}{T}\right)^{-2n} \frac{dT}{T} \quad \text{and} \quad K(V) \exp\left\{-V\left(\frac{T}{\theta} + \frac{\theta}{T}\right)\right\} \frac{dT}{T},$$

where $K$ and $K(V)$ are independent of $T$. The two distributions may be used to give respectively unconditional and conditional systems of confidence intervals for $\theta$. To use the former, put $\beta = T/\theta$ and determine constants $\beta_1$ and $\beta_2$ such that

$$\int_{\beta_1}^{\beta_2} \beta^{2n-1}(1+\beta^2)^{-2n}\, d\beta = \alpha \int_0^\infty \beta^{2n-1}(1+\beta^2)^{-2n}\, d\beta.$$

This leads to the assertion

$$\text{“}P(A) = P(\beta_1 \leqq T/\theta \leqq \beta_2) = P(T/\beta_2 \leqq \theta \leqq T/\beta_1) = \alpha\text{”}.$$

But conditionally upon the value of the ancillary $V$ the probability is

$$P(A \mid V) = \int_{\beta_1\theta}^{\beta_2\theta} \exp\left\{-V\left(\frac{T}{\theta}+\frac{\theta}{T}\right)\right\}\frac{dT}{T} \Big/ \int_0^\infty \exp\left\{-V\left(\frac{T}{\theta}+\frac{\theta}{T}\right)\right\}\frac{dT}{T}$$

$$= \int_{\beta_1}^{\beta_2} \exp\left\{-V\left(\beta+\frac{1}{\beta}\right)\right\}\frac{d\beta}{\beta} \Big/ \int_0^\infty \exp\left\{-V\left(\beta+\frac{1}{\beta}\right)\right\}\frac{d\beta}{\beta}$$

The last expression depends on the ancillary $V$ but is independent of the parameter $\theta$; it will equal $\alpha$ only for a particular intermediate value of $V$, and for all other values relevant subsets will be defined.

3.4. *Another example involving an ancillary statistic.* Turning to Fisher ([9], and [11], p. 134) for another example involving an ancillary statistic, we let $x$ and $y$ have a circular normal distribution with unit variance and with mean on a circle of known radius $R$ so that the density is

$$(1/2\pi)\exp\{-\tfrac{1}{2}(x - R\cos\theta)^2 - \tfrac{1}{2}(y - R\sin\theta)^2\}\, dx\, dy$$

where $\theta$ is to be estimated. If a single observation $(x, y)$ is expressed in polar coordinates by

$$a^2 = x^2 + y^2 \qquad x = a\cos\hat{\theta}$$
$$\hat{\theta} = \tan^{-1} y/x \qquad y = a\sin\hat{\theta}$$

then $\hat{\theta}$ is the maximum likelihood estimator of $\theta$, and $a$ is ancillary. One easily obtains the joint density

$$(a/2\pi)\exp\{-\tfrac{1}{2}(a^2 + R^2 - 2aR\cos(\hat{\theta} - \theta))\}\, da\, d\hat{\theta}.$$

An unconditional system of confidence intervals is given by

$$\text{“}P(\hat{\theta} - \gamma \leqq \theta \leqq \hat{\theta} + \gamma) = \alpha\text{”}$$

where $\alpha$ is the fraction of the unit circular normal distribution inside a wedge of angle $2\gamma$, the maximum density being centered in the wedge at a distance $R$ from the vertex. To find the probability conditioned on the ancillary, we find the marginal distribution of $a$ to be expressible in terms of the Bessel function $I_0$ by

$$\frac{a\, da}{2\pi}\exp\{-\tfrac{1}{2}(a^2 + R^2)\}\int_0^{2\pi} \exp\{aR\cos(\hat{\theta} - \theta)\}\, d\hat{\theta}$$
$$= a\exp\{-\tfrac{1}{2}(a^2 + R^2)\} I_0(aR)\, da.$$

Thus the conditional density of $\hat{\theta}$ given $a$ is

$$\frac{\exp\{aR\cos(\hat{\theta}-\theta)\}}{2\pi I_0(aR)}\,d\hat{\theta},$$

and the conditional probability for a given value of $a$ is

$$P(\hat{\theta}-\gamma \leqq \theta \leqq \hat{\theta}+\gamma \mid a) = \frac{1}{2\pi I_0(aR)}\int_{-\gamma}^{\gamma} e^{aR\cos\phi}\,d\phi.$$

As in the previous example, the conditional probability depends on the ancillary and is independent of the parameter. Here it approaches unity as $a$ tends to infinity and approaches $\gamma/\pi$ as $a$ tends to zero. Thus subsets constructed from observations lying near the origin will be negatively biased; the variability of $\hat{\theta}$ for these considered separately is larger than the variability of $\hat{\theta}$ for all observations collectively.

3.5. *An example involving the Behrens-Fisher problem.* Behrens' hypothesis states that the means $\mu_1$, $\mu_2$ of two normal populations differ by $\delta(=0$, usually), no assumption of the equality of variances $\sigma_1^2$, $\sigma_2^2$ being made. We consider the test devised by Welch [21] tabulated by Aspin [1] and appearing as Table 11 in the *Biometrika* Tables of Pearson and Hartley [18]. This test rejects the null hypothesis $H_0$: $\mu_1 - \mu_2 = \delta$ when $|\bar{x}_1 - \bar{x}_2 - \delta| > v(s_1^2/n_1 + s_2^2/n_2)^{1/2}$ where $n$, $\bar{x}$, $s^2$ denote size, mean, and variance of the samples and $v$ (the analog of Student's $t$) is a tabulated function of $n_1$, $n_2$, $s_1^2$, $s_2^2$, and the significance level. A calculation of Fisher [10] shows that for fixed $s_2/s_1$ the conditional probability $P$ (reject $\mid H_0$, $s_2/s_1$) can be expressed as a unique function of the ratio $\sigma_2^2/\sigma_1^2$; and that when $n_1 = n_2 = 7$, $s_1 = s_2$, and the significance level is 0.1, then $P$ (reject $\mid H_0$, $s_1 = s_2$) achieves a *minimum* value[2] of 0.108 when $\sigma_2^2/\sigma_1^2 = 1$. Now if the tabulated test is translated into a system of confidence intervals, one obtains the probability assertion

$$\text{“}P(A) = P\{|(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)| \leqq v(s_1^2/n_1 + s_2^2/n_2)^{1/2}\} = \alpha\text{”}$$

for which Fisher's calculation shows that when $n_1 = n_2 = 7$, $\alpha = 0.9$,

$$P(A \mid C) \leqq \alpha - \epsilon \quad \text{for all} \quad \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$$

where $\epsilon = 0.008$ and $C$ is the subset $s_1 = s_2$.

Thus the tabulated solution induces a negatively biased relevant subset. It is noteworthy that in this example $P(A \mid C)$ depends on the ratio $\sigma_2^2/\sigma_1^2$ whereas in all four preceding examples $P(A \mid C)$ was independent of the parameters. One might well wish to distinguish between these two types of relevance, although separate names have not been advanced here.

In [10] Fisher seems to imply that subsets defined by fixed values of $s_2/s_1$ are uniquely appropriate reference sets for inferences about Behrens' hypothesis; but like Bartlett [2] and Welch [22], I do not find his reasons to be compelling. Fisher does show that the subset $s_1 = s_2$ is not relevant (in the present technical

[2] Fisher gives a graph but no table. The value 0.108 is obtained from Welch [22].

sense) for Behrens' solution, but he does not consider the possible relevance of other subsets. Wallace [20] has since shown that no such relevant subsets exist.

3.6. *An example in which an individual point of the sample space is a relevant subset.* As a rule the property of being relevant belongs to a set of possible observations and not to individual points of the observation space. But in rare cases individual points may be relevant, as can be illustrated by a familiar component-of-variance problem. Let

$$y_{ij} = \mu + a_i + e_{ij} \quad (i = 1, \cdots, r; \; j = 1, \cdots, s)$$

in which $\mu$ is an unknown parameter, $a_i$ and $e_{ij}$ are normal and independently distributed with zero means and variances $\sigma_a^2$ and $\sigma^2$. Using the dot notation to indicate an average over the missing subscript, the sums of squares

$$Q_1 = \sum\sum (y_{ij} - y_{i\cdot})^2 \quad \text{and} \quad Q_2 = \sum\sum (y_{i\cdot} - y_{\cdot\cdot})^2$$

are known to have independent distributions given in terms of chi-square by

$$\sigma^2 \chi^2_{r(s-1)} \quad \text{and} \quad (\sigma^2 + s\sigma_a^2)\chi^2_{r-1}.$$

Thus the quantity

$$\frac{(\sigma^2 + s\sigma_a^2)\, f_2\, Q_1}{\sigma^2 f_1\, Q_2} = \frac{\chi^2_{f_1}/f_1}{\chi^2_{f_2}/f_2}$$

has the $F$ distribution with $f_1 = r(s - 1)$ and $f_2 = r - 1$ degrees of freedom. If lower and upper percentage points $F_1$ and $F_2$ are chosen so that

$$P(F_1 \leqq F \leqq F_2) = \alpha$$

then on rearranging the inequalities one finds that the probability is $\alpha$ that

$$\frac{1}{s}\left\{\frac{f_1\, Q_2}{f_2\, Q_1} F_1 - 1\right\} \leqq \frac{\sigma_a^2}{\sigma^2} \leqq \frac{1}{s}\left\{\frac{f_1\, Q_2}{f_2\, Q_1} F_2 - 1\right\}.$$

If the ratio of mean squares is larger than the upper percentage point,

$$\frac{Q_1/f_1}{Q_2/f_2} > F_2$$

as always can happen with at least some small probability, then the confidence interval includes only negative values. The probability of covering the true ratio $\sigma_a^2/\sigma^2$ conditionally on any such observation is zero; thus such individual points constitute negatively biased relevant subsets.

Now let $C$ denote the collection of all such points, that is, all observations for which the last inequality is satisfied. Then $C$ is relevant and negatively biased. Is increased confidence thereby justified for observations in the complementary set $C'$? That $C'$ is *semirelevant* follows from Lemma 3. Using $P(A \mid C) = 0$ one obtains

$$P(A \mid C') = \alpha/P(C') = \alpha + \alpha P(C)/P(C')$$

so that the bias is $\alpha P(C)/P(C')$. Since $P(C)$ tends to zero as $\sigma^2/\sigma_a^2$ tends to zero, there is no positive lower bound for the bias. Thus the complementary set $C'$ is not relevant. Loosely summarizing these results: within $C$ one's confidence is necessarily zero; within $C'$, $\alpha$ is the greatest lower bound for the confidence over the possible states of nature.

**4. Examples of semirelevant subsets.** From the following examples it will be seen that semirelevant subsets are more easily found than relevant ones and that the requirement that none should exist is a very severe restriction indeed.

4.1. *A nonparametric assertion about the median.* Let $g(t)$ be any continuous density which is positive for $-\infty < t < \infty$ and which has median equal to zero. If one observation $x$ is taken from the density $g(x - \theta)$ ($\theta$ = median), then the assertion

$$\text{“}P(A) = P(x < \theta < \infty) = \tfrac{1}{2}\text{”}$$

is weakly exact. Consider the subset $C$ defined by $x > 0$. The conditional probability $P(A \mid C) = P(x < \theta \mid x > 0) = P(0 < x < \theta)$ is zero for negative $\theta$ while for positive $\theta$ one has

$$P(0 < x < \theta) = \int_0^\theta g(x - \theta)\, dx = \int_{-\theta}^0 g(t)\, dt$$

$$= \tfrac{1}{2} - \int_{-\infty}^{-\theta} g(t)\, dt < \tfrac{1}{2}$$

for all $0 < \theta < \infty$. Thus $C$ is semirelevant, negatively biased. The last integral, which gives the bias as a function of $\theta$, is seen to tend to zero as $\theta$ tends to infinity; thus $C$ is semirelevant only—not relevant.

4.2. *Student's t.* Let $x$ be normally distributed with unknown mean $\mu$ ($-\infty < \mu < \infty$) and unknown variance $\sigma^2$ ($0 < \sigma^2 < \infty$). If $\bar{x}$ and $s^2$ denote the sample mean and variance, then the conventional confidence or fiducial interval estimate of $\mu$ corresponds to the assertion

$$\text{“}P(\bar{x} - ks \leqq \mu \leqq \bar{x} + ks) = \alpha\text{”}$$

where $k = t_{\alpha,n-1}/\sqrt{n}$ and $t_{\alpha,n-1}$ is the appropriate percentage point of Student's $t$. We shall first show that for any $a > 0$, the subset $s < a$ is semirelevant, negatively biased. Stated in another way, for any $\mu$, $\sigma^2$, "long" intervals cover the true value more frequently than "short" intervals, and this fact holds for any critical length which is used to distinguish "long" intervals from "short" ones.

Denote by $A$ the event $|\bar{x} - \mu| \leqq ks$, by $C$ and $C'$ the subset $s < a$ and its complement $s \geqq a$, by $\sigma^{-1}f(s/\sigma)\, ds$ and $g(\bar{x} - \mu)\, dx$ the independent densities of $s$ and $\bar{x}$. Then

$$P(C) = \int_0^a \sigma^{-1} f(s/\sigma)\, ds = \lambda(a/\sigma), \qquad \text{say,}$$

and

$$P(AC) = \int_0^a \sigma^{-1} f(s/\sigma) \int_{\mu-ks}^{\mu+ks} g(\bar{x} - \mu)\, d\bar{x}\, ds.$$

If we put

$$G(s) = \int_{\mu-ks}^{\mu+ks} g(\bar{x} - \mu)\, d\bar{x} = \int_{-ks}^{ks} g(z)\, dz,$$

then by a mean value theorem

$$P(AC) = \int_0^a \sigma^{-1} f(s/\sigma) G(s)\, ds = G(s_0) \int_0^a \sigma^{-1} f(s/\sigma)\, ds = G(s_0)\lambda(a/\sigma)$$

where $0 < s_0 < a$. This gives

$$P(A \mid C) = P(AC)/P(C) = G(s_0).$$

By a similar argument

$$P(A \mid C') = G(s_0')$$

where $a < s_0' < \infty$. But $G$ is an increasing function so that $s_0 < a < s_0'$ implies $G(s_0) < G(s_0')$. The values of $s_0$ and $s_0'$ depend on $a$ and $\sigma$, but for any particular $a$, $\sigma$, weak exactness implies that $C$ and $C'$ must have opposite bias. Thus for all $a$, $\sigma$, $\mu$,

$$G(s_0) = P(A \mid C) < \alpha < P(A \mid C') = G(s_0')$$

so that $C$ and $C'$ are semirelevant.

To illustrate the situation more clearly, we consider the simple case $n = 2$, $\alpha = \frac{1}{2}$, for which $s^2 = \frac{1}{2}(x_1 - x_2)^2$ and the assertion is that the mean has an even chance of lying between the two observations:

$$\text{“}P(A) = P(x_{\min} \leqq \mu \leqq x_{\max}) = \tfrac{1}{2}\text{”}.$$

For the subset $C'$: $s \geqq a$, the conditional probability is

$$P(A \mid C') = \tfrac{1}{2} + \tfrac{1}{2}\phi(a/\sigma)$$

where $\phi(a/\sigma)$ is the standard normal probability between $-a/\sigma$ and $+a/\sigma$. Thus the bias, $\frac{1}{2}\phi(a/\sigma)$, is always positive, increasing from 0 to $\frac{1}{2}$ as $a$ increases from 0 to $\infty$. If Paul adopts the strategy: bet even odds that $A$ is true when $s \geqq a$ and bet even odds that $A$ is false when $s < a$, then the probability that Paul wins is

$$P = \tfrac{1}{2} + \phi(1 - \phi), \qquad \phi = \phi(a/\sigma),$$

which lies in the range $1/2 < P \leqq 3/4$, and Paul's expected gain on these bets of 1/2 unit is

$$G = \phi(1 - \phi)$$

which always lies in the range $0 < G \leqq 1/4$. The maxima of $P$ and $G$ are at-

tained when $\phi = 1/2$, that is, when

$$a = 0.67\sigma, \quad \text{approximately.}$$

Thus the optimum value of $a$ is proportional to $\sigma$; and for dimensional reasons this is true generally, the proportionality factor depending on $n$ and $\alpha$.

In an idealized model Paul will have no prior information concerning $\sigma$ and thus will have no basis for a choice of $a$ in the range from zero to infinity, but in practically any applied problem some prior knowledge of $\sigma$ will be available. It is noteworthy that Fisher has specifically stated that fiducial inference is valid only in the absence of prior knowledge ([11], p. 51). In contrast, Neyman appears not to have taken a stand on the applicability of confidence interval theory in the presence of prior information. I believe that the above calculations tend to justify Fisher's restriction. Now the assumption that a prior distribution is known and the assumption that *nothing* is known are two boundaries of a vast intermediate area in which there is *partial* prior information. We have seen how Paul can find a crude stategy in this middle area against Peter's conventional solution. But what one would really like to know is how Peter's rule might be altered to use partial prior knowledge. That is a much more subtle and difficult question.

**5. An example of nonexistence of relevant subsets.** Let $g(t)$ be a continuous density which is nonzero for all $t$, and let $f(x, \theta) = g(x - \theta)$ so that $\theta$ is a location parameter. From a sample of one value of $x$, $\theta$ may be estimated by the assertion

$$\text{“}P(A) = P(-t_1 \leqq x - \theta \leqq t_2) = P(x - t_2 \leqq \theta \leqq x + t_1) = \alpha\text{”}$$

where $t_1$ and $t_2$ are chosen so that

$$\int_{-t_1}^{t_2} g(t)\, dt = \alpha.$$

We wish to show that neither of the inequalities

$$(1) \qquad P(A \mid C) \geqq \alpha + \epsilon \quad \text{or} \quad P(A \mid C) \leqq \alpha - \epsilon \qquad \text{(for all } \theta)$$

can hold for any Lebesgue measurable set $C$ of values of $x$ having finite or infinite, but not zero, Lebesgue measure (one is prevented from treating sets of measure zero by the nonuniqueness of the definition of conditional probability; see for example [13], p. 12). Let $a(\theta)$, $b(\theta)$, $A(R)$, $B(R)$ be defined by

$$a(\theta) = P(C), \qquad A(R) = \int_{-R}^{R} a(\theta)\, d\theta,$$

$$b(\theta) = P(AC), \qquad B(R) = \int_{-R}^{R} b(\theta)\, d\theta.$$

Then $P(A \mid C) = b(\theta)/a(\theta)$; and substituting in (1), multiplying by $a(\theta)$,

integrating over $-R \leqq \theta \leqq R$, and dividing by $A(R)$ gives

$$\frac{B(R)}{A(R)} \geqq \alpha + \epsilon \text{ or } \frac{B(R)}{A(R)} \leqq \alpha - \epsilon \quad (\text{for all } R).$$

It is shown in the appendix that subject to weak conditions on $g(t)$, $B(R)/A(R) \to \alpha$ as $R \to \infty$, and thus the desired result is established by contradiction. Some sweeping generalizations of this result have been found by Wallace [20].

**6. Generalization of the model.** It may first be noted that the theory of *tolerance* as well as confidence intervals may be treated simply by taking $A$ to be the proposition that at least a certain proportion of the population lies between stated limits.

A useful generalization consists in representing the outcome of the random experiment by a pair of random variables $(x, y)$ of which $x$ is known and $y$ is unknown to the players, both being known to the referee. The theory of *prediction intervals* is then treated by taking $x$ to be "past" observations and $y$ to be "future" observations to be predicted (it is immaterial that the "future" observations $y$ have already been observed by the referee so long as they are unknown to the two players). The proposition $A$ is then a statement depending on $x$ concerning the future observations; a notable distinction is that $A$ does not concern the unknown state of nature $U$.

In *Bayesian interval estimation* (Cramér [6], p. 508, Neyman [17], p. 162) $y$ plays the role of the unknown parameter having a known prior distribution. Bayesian problems generally have the distinguishing feature that the state of nature is assumed known so that $U$ does not appear. As in prediction interval theory, the proposition $A$ gives a relation between $x$ and $y$, the known and unknown observations.

A further possible generalization which might have some interest, although it seems not to be required for existing theories, consists in allowing the confidence coefficient $\alpha$ to depend on the known observations: $\alpha = \alpha(x)$. That is, the rule $R$ for probability assertions may specify not only how the proposition $A$ is to depend on $x$ but also how the asserted probability level is to depend on $x$. The definition of relevant subsets then requires some repair; the logical extension consists in considering only subsets $C$ of values of $x$ for which $\alpha(x)$ takes a fixed value. An example of variable $\alpha$ is given in Section 9.

6.1. *The criterion of strong exactness.* It has been noted that the expected gain to Paul is a function of the unknown state of nature $U$, Peter's rule $R$, and Paul's strategy $S$; thus it may be denoted by $G(U, R, S)$.

*Definition.* A rule $R_0$ will be called *strongly exact* if

$$G(U, R_0, S) = 0 \quad \text{identically for all } U, S.$$

In other words: Whatever the true state of nature and whatever strategy Paul may use, the expected gain to Paul is zero. It is essentially equivalent to write $P(A \mid C) = \alpha$ for all $U, C$.

It appears to be impossible to satisfy the very stringent condition of strong exactness except in rather special cases, e.g., in Bayesian estimation where the condition "all $U$" is in fact no requirement at all since $U$ is not variable. Thus strong exactness is not so much a practical requirement as a goal toward which one might strive even though it cannot actually be reached.

**7. Remarks on Bayesian interval estimation.** In a typical Bayesian situation an experiment consists in obtaining a random value of the "parameter" from a known distribution $w(\theta)$ and subsequently observing values $x_1, \cdots, x_n$ from a distribution $f(x; \theta)$ which is known but for the value of $\theta$. Thus in the model of Section 6, $x$ represents $x_1, \cdots, x_n$ and $y$ represents $\theta$. The conditional distribution of $\theta$ given $x_1, \cdots, x_n$ is

$$h(\theta \mid x_1, \cdots, x_n) = \frac{w(\theta)\{f(x_1;\theta) \cdots f(x_r;\theta)\}}{\int w(\theta)\{f(x_1;\theta) \cdots f(x_n;\theta)\}\, d\theta}$$

A Bayesian estimate of $\theta$ is given by the assertion

$$\text{“}P(k_1 \leqq \theta \leqq k_2) = \alpha\text{”}$$

where $k_1$ and $k_2$ are any two numbers depending on the observations by

$$\int_{k_1}^{k_2} h(\theta \mid x_1, \cdots, x_n)\, d\theta = \alpha$$

Apart from the arbitrary allocation of the probability $1 - \alpha$ between the two tails of the distribution, $k_1$ and $k_2$ are unique functions of the observations. It may further be noted that the interval $(k_1, k_2)$ based on any particular observation $(x_1, \cdots, x_n)$ has its own validity without reference to intervals which might be defined for other observations; this is in contrast to confidence interval theory in which any particular interval is meaningful only when referred to a system of intervals defined for all possible samples. A closely connected fact is that the above Bayesian solution gives a rule for assertions which is strongly exact.

Two variations of the above are found in the literature. Cramér ([6], p. 508) replaces $h(\theta \mid x_1, \cdots, x_n)$ by

$$h(\theta \mid \hat{\theta}) = \frac{w(\theta)g(\hat{\theta};\theta)}{\int w(\theta)g(\hat{\theta};\theta)\, d\theta}$$

where $\hat{\theta}$ is an estimator of $\theta$ and $g$ is the density of $\hat{\theta}$. Three remarks can be made: (i) If $\hat{\theta}$ is sufficient for $\theta$ and if the allocation of the probability $1 - \alpha$ to the two tails is fixed, then this solution will not differ from the preceding one. (ii) If $\hat{\theta}$ is not sufficient and if the sample $(x_1, \cdots, x_n)$ is known to the players, then relevant subsets will exist and the solution is weakly exact but not strongly exact. (iii) If $\hat{\theta}$ is not sufficient, but the players have knowledge only of $\hat{\theta}$ and not of $(x_1, \cdots, x_n)$, then the last solution is strongly exact.

A second variation is given by Neyman's "modernized Bayes' estimating intervals" ([17], pp. 165–181) in which the expected length of the intervals is shortened by requiring that the asserted probability $\alpha$ refer only to the relative frequency in the sequence of all experiments, not separately to sequences of fixed $(x_1, \cdots, x_n)$. Thus in the "modernized Bayes' solution" we have a clear-cut example of a rule that is weakly but not strongly exact. This situation is quite similar to the Cox example of Section 3.2.

**8. A familiar prediction example.** To show how a prediction problem fits into the general scheme and to point out some analogies between prediction and confidence intervals, we consider a familiar prediction example. Let $x$ have a continuous density of unspecified form, let $x_1$ and $x_2$ represent two known observations, and let $x_3$ represent an unknown or future observation. Then since all six of the permutations of $x_1 < x_2 < x_3$ are equally likely, the probability assertion

$$\text{“}P(x_{\min} < x_3 < x_{\max}) = \tfrac{1}{3}\text{”}$$

is weakly exact. Some years ago this particular example figured in a dispute between Fisher [7], [8] and Jeffreys [15], [16]. Without claiming to understand the subtleties of the dispute we note that Fisher [7] points out that "for any particular population the probability will generally be larger when the first two observations are far apart than when they are near together." From Fisher's remark it follows that just as in the Student's $t$ example of Section 4.2, long intervals are valid more often than short ones, and the subsets $x_{\max} - x_{\min} \leqq a$ or $\geqq a$ are semirelevant for any $a > 0$. Fisher goes on to say that "the fallacy of Jeffreys' argument consists just in assuming that the probability shall be 1/3, *independently of the distance apart of the first two observations*." In the present terminology we would say that Jeffreys is accused of treating an assertion as if it were strongly exact when in fact it is only weakly exact.

**9. An example of intersecting relevant subsets.** In this section some rather artificial examples are constructed primarily to illustrate intersecting relevant subsets. The examples differ from those preceding in that the proposition $A$ is made independent of the observations and the "confidence level" $\alpha$ is allowed to be random. The examples raise the question of whether interesecting relevant subsets might exist elsewhere, for example, in confidence interval situations.

Let an experiment consist in drawing one ball from twelve contained in an urn. Let $A$, $B$, $D$ denote attributes, and let $A'$, $B'$, $D'$ denote the respective negations. We suppose that Peter and Paul have knowledge of the contents of the urn as shown in Table 1, and that the referee draws one ball and announces whether $B$ or $B'$ is observed ($x = B$ or $B'$ = observation known to players). It is not revealed whether $A$ or $A'$ is observed ($y = A$ or $A'$ = unknown observation). Peter's probability assertion concerns the probability that the ball drawn has attribute $A$. The assertion "$P(A) = 5/12$" is weakly exact. It is objectionable on the grounds that the (only) subsets $B$ and $B'$ are relevant.

A strongly exact assertion clearly can be made by allowing a variable "confidence level":

$$\text{“}P(A) = \begin{cases} 4/9 & \text{if } x = B \\ 1/3 & \text{if } x = B' \end{cases}\text{”}$$

The complexion of the problem is changed if the players are given further information. Suppose the twelve balls are also classified in categories $D$ and $D'$ and that the values in Table 2 are also known to the players. The referee announces one of four possible results: $x = BD, BD', B'D$, or $B'D'$. It will be seen that no probability assertion of the form "$P(A) = \alpha(x)$" is strongly exact inasmuch as the values of $P(A \mid BD)$, etc., are dependent on an unknown variable state of nature, for the two *marginal* $2 \times 2$ tables fail to specify a unique $2 \times 2 \times 2$ table which would describe the contents completely. Table 3 shows that there are six possible states of nature which may be obtained by putting $a = 2, 3$ (= number of $ABD$) and $a' = 1, 2, 3$ (= number of $A'BD$).

Table 1

| | $A$ | $A'$ | |
|---|---|---|---|
| $B$ | 4 | 5 | 9 |
| $B'$ | 1 | 2 | 3 |
| | 5 | 7 | 12 |

Table 2

| | $A$ | $A'$ | |
|---|---|---|---|
| $D$ | 3 | 3 | 6 |
| $D'$ | 2 | 4 | 6 |
| | 5 | 7 | 12 |

Table 3

| | $A$ | | | $A'$ | | |
|---|---|---|---|---|---|---|
| | $B$ | $B'$ | | $B$ | $B'$ | |
| $D$ | $a$ | $3 - a$ | 3 | $a'$ | $3 - a'$ | 3 |
| $D'$ | $4 - a$ | $a - 2$ | 2 | $5 - a'$ | $a' - 1$ | 4 |
| | 4 | 1 | 5 | 5 | 2 | 7 |

Intersecting relevant subsets are obtained from the weakly exact probability assertion, "$P(A) = 5/12$," for which $B$ and $D$ are relevant, positively biased and $B'$ and $D'$ are relevant and negatively biased. $BD$ and $B'D'$ are intersections of similarly biased subsets; $BD'$ and $B'D$ are intersections of dissimilarly biased subsets. It is interesting to note that $BD$ itself is *not* relevant, for with $a = 2$, $a' = 3$ one has $P(A \mid BD) = 2/5 < 5/12$. Thus Paul has a positive expectation if he bets $A$ is true when $B$ is observed, or if he bets $A$ is true when $D$ is observed; but he may have a negative expected gain if he bets $A$ is true only when *both* are observed!

Is there a uniquely appropriate reference set for Peter's assertions in the last example? Use of the universal set ignores information about $B$ and $D$. Use of $BD$, $BD'$, $B'D$, and $B'D'$ seems inappropriate because the relative frequency of $A$ is not known within these subsets. It would seem about equally as appropriate to use $B$ and $B'$ as to use $D$ and $D'$. Thus there appear to be no uniquely appropriate subsets.

**10. Summary and conclusions.** Statistical inferences having the form, "The probability that $A$ is true is equal to $\alpha$" can be studied within the framework of a game between two players, one who makes such inferences (or *probability assertions*) and an opponent who questions their validity. The model suggests

a number of criteria of validity of such inferences; four criteria which have been defined and illustrated by examples are: (i) weak exactness, (ii) strong exactness, (iii) nonexistence of relevant subsets, and (iv) nonexistence of semirelevant subsets. Definitions of these concepts, too lengthy to be repeated here, are given in Sections 2.1, 6.1, 2.2, and 2.2, respectively. Some general observations are:

(i) Weak exactness is a criterion suggested by a familiar requirement of Neyman's in the theory of confidence intervals; in the general model of Section 6 the definition is extended to apply to more general problems.

(ii) Strong exactness is a much more severe requirement which can be satisfied in classical Bayesian estimation situations, but which appears to be unreasonably demanding in most non-Bayesian problems. Those who eschew the prior distributions of the classical theory pay for weakening the classical assumptions by losing the property of strong exactness of the inferences. To mistake weak exactness for strong exactness is to attribute to an inference a more desirable property than it actually possesses. The logical fallacy is neatly stated by Sir Macklin [14]:

"Then I shall demonstrate to you,
According to the rules of Whately,
That what is true of all, is true
Of each, considered separately."

(iii) The criterion of nonexistence of relevant subsets is largely inspired by some recent work of Fisher. Various examples of relevant subsets have been given in order to provide a better understanding of their nature. Nonexistence is established only for one simple case; for much more general results the reader is referred to Wallace [20].

(iv) From the examples of Section 4 it is seen that nonexistence of semirelevant subsets is a very severe requirement indeed. One may conjecture that fiducial intervals do not induce *relevant* subsets, but from the example of Student's $t$ it is seen that the same conjecture is not true for *semirelevant* subsets.

It is to be hoped that eventually there will be found some generally accepted notion of an "appropriate reference set" for inferences. Some readers may find that the examples of Sections 3.2, 3.3, and 3.4 indicate that the universal set is not always as appropriate as some suitably chosen subset. On this subject, Fisher ([11], p. 110) states:

> "If, therefore, any portion of the data were to allow of the recognition of such a subset, to which the predicand belongs, a different probability would be asserted using the smallest such subset recognizable."

Perhaps Fisher's idea should be formulated mathematically in terms of minimal subsets on which probabilities are known independently of the unknown state of nature. That uniqueness of appropriate reference sets might be a problem is indicated by the example of Section 9, in which two different subdivisions of the observation space give different reference sets which are about equal in merit.

The appropriate reference set has been a subject of controversy in testing situations (i.e., tests of significance and tests of hypothesis); contingency tables and regression problems (Fisher [11], pp. 82–88) are old examples. Some recent examples can be found in Cox [5] and in Cohen [3]. It is to be noted that the present development has been based largely on problems of interval *estimation*. The usual translation of criteria to *testing* situations is of course possible in many cases. Thus certain testing situations have been treated implicitly in this work; but perhaps the reader will find that whatever force the arguments may have for estimation problems is diminished in the translation to testing problems.

**11. Acknowledgement.** I wish to thank Prof. Wallace for giving me a draft of his paper prior to publication.

## APPENDIX

*Proof of the result of Section 5*

To establish the result of Section 5 it will be assumed that the cumulative distribution $G(t) = \int_{-\infty}^{t} g(t)\, dt$ satisfies the mild restrictions

$$\int_{-\infty}^{0} G(t)\, dt < \infty \quad \text{and} \quad \int_{0}^{\infty} (1 - G(t))\, dt < \infty .$$

Denote by $\phi_C(x)$ the set characteristic function of the set $C$. Then

$$a(\theta) = P(C) = \int_{-\infty}^{\infty} \phi_C(x) g(x - \theta)\, dx$$

$$b(\theta) = P(AC) = \int_{t_1+\theta}^{t_2+\theta} \phi_C(x) g(x - \theta)\, dx$$

Further define $\mu(R)$ by

$$\mu(R) = \int_{-R}^{R} \phi_C(x)\, dx$$

Then if $\mu(\infty)$ is finite, a straightforward calculation shows that $A(R) \to \mu(\infty)$ and $B(R) \to \alpha\mu(\infty)$ as $R \to \infty$ which establishes the desired contradiction.

If $\mu(R) \to \infty$, we may compare $A(R)$ with

$$\begin{aligned} A'(R) &= \int_{\theta=-\infty}^{\infty} \int_{x=-R}^{R} \phi_C(x) g(x - \theta)\, dx\, d\theta \\ &= \int_{x=-R}^{R} \phi_C(x) \int_{\theta=-\infty}^{\infty} g(x - \theta)\, d\theta\, dx \\ &= \int_{-R}^{R} \phi_C(x)\, dx = \mu(R) . \end{aligned}$$

The difference $A(R) - A'(R)$ is made up of four integrals described (ignoring signs) by: $x$ (or $\theta$) in the range $-R$ to $R$, and $\theta$ (or $x$) in the range $\pm R$ to $\pm\infty$.

All four are bounded by virtue of the assumption on $G$. A typical calculation follows:

$$\int_{x=-R}^{R}\int_{\theta=R}^{\infty}\phi_C(x)g(x-\theta)\,d\theta\,dx \leq \int_{x=-R}^{R}\int_{\theta=R}^{\infty}g(x-\theta)\,d\theta\,dx$$
$$= \int_{x=-R}^{R}\int_{t=-\infty}^{x-R}g(t)\,dt\,dx$$
$$= \int_{-R}^{R}G(x-R)\,dx$$
$$= \int_{-2R}^{0}G(x')\,dx'$$

Thus $A(R) = \mu(R) + K(R)$ where $|K(R)|$ is bounded as $R$ tends to infinity. The expression for $B(R)$ may be treated as follows:

$$B(R) = \int_{\theta=-R}^{R}\int_{t=-t_1}^{t_2}\phi_C(t+\theta)g(t)\,dt\,d\theta$$
$$= \int_{t=-t_1}^{t_2}\int_{\theta=-R}^{R}\phi_C(t+\theta)g(t)\,d\theta\,dt$$
$$= \int_{t=-t_1}^{t_2}\int_{x=t-R}^{t+R}\phi_C(x)g(t)\,dx\,dt$$
$$= K'(R) + \int_{t=-t_1}^{t_2}\int_{x=-R}^{R}\phi_C(x)g(t)\,dx\,dt$$
$$= K'(R) + \alpha\mu(R)$$

where $K'(R)$ is the sum of two positive and two negative terms, each term being an integral over a region whose area is independent of $R$. Thus $|K'(R)|$ is bounded. Consequently the ratio

$$\frac{B(R)}{A(R)} = \frac{\alpha\mu(R) + K'(R)}{\mu(R) + K(R)}$$

tends to $\alpha$ and $R$ tends to infinity, and the contradiction is established.

## REFERENCES

[1] A. A. Aspin, "Tables for use in comparisons whose accuracy involves two variances, separately estimated," *Biometrika*, Vol. 36 (1949), pp. 290–293.

[2] M. S. Bartlett, "Comment on Sir Ronald Fisher's paper: 'On a test of significance in Pearson's Biometrika Tables (No. 11),' " *J. Roy. Stat. Soc., Ser. B*, Vol. 18 (1956), pp. 295–296.

[3] Leonard Cohen, "On mixed single sample experiments," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 947–971.

[4] D. R. Cox, "Some problems connected with statistical inference," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 357–372.

[5] D. R. Cox, "The regression analysis of binary sequences," *J. Roy. Stat. Soc., Ser. B*, Vol. 20 (1958), pp. 215–242.

[6] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[7] R. A. Fisher, "The concepts of inverse probability and fiducial probability referring to unknown parameters," *Proc. Roy. Soc. London, Ser. A*, Vol. 139 (1933), pp. 343–348.

[8] R. A. Fisher, "Probability, likelihood, and the quantity of information in the logic of uncertain inference," *Proc. Roy. Soc. London, Ser. A*, Vol. 146 (1934), pp. 1–8.

[9] Sir Ronald Fisher, "Statistical methods and scientific induction," *J. Roy. Stat. Soc., Ser. B*, Vol. 17 (1955), pp. 69–78.

[10] Sir Ronald Fisher, "On a test of significance in Pearson's Biometrika Tables (No. 11)," *J. Roy. Stat. Soc., Ser. B*, Vol. 18 (1956), pp. 56–60.

[11] Sir Ronald A. Fisher, *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh, 1956.

[12] Sir Ronald A. Fisher, "Mathematical probability in the natural sciences," *Technometrics*, Vol. 1 (1959), pp. 21–29.

[13] D. A. S. Fraser, *Nonparametric Methods in Statistics*, Wiley, New York, 1957.

[14] W. S. Gilbert, "Sir Macklin" in *The Bab Ballads*, Macmillan, London, 1898.

[15] H. Jeffreys, "On the theory of least squares," *Proc. Roy. Soc. London, Ser. A*, Vol. 138 (1932), pp. 48–55.

[16] H. Jeffreys, "Probability, statistics, and the theory of errors," *Proc. Roy. Soc. London, Ser. A*, Vol. 140 (1933), pp. 523–535.

[17] Jerzy Neyman, *Lectures and Conferences on Mathematical Statistics and Probability*, U. S. Department of Agriculture, Washington, D. C., 1952.

[18] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians, Vol. I*, Cambridge University Press, 1954.

[19] Richard von Mises, *Probability, Statistics and Truth*, Allen and Unwin, London, 1957.

[20] David L. Wallace, "Conditional confidence level properties," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 864–876.

[21] B. L. Welch, "The generalization of 'Student's' problem when several different population variances are involved," *Biometrika*, Vol. 34 (1947), pp. 28–35.

[22] B. L. Welch, "Note on some criticisms made by Sir Ronald Fisher," *J. Roy. Stat. Soc., Ser. B*, Vol. 18 (1956), pp. 297–302.

# CONDITIONAL CONFIDENCE LEVEL PROPERTIES[1]

By David L. Wallace

*University of Chicago*

**1. Introduction.** Some confidence region procedures have the property that, conditionally on the sample point lying in some subset of the sample space, the conditional confidence level (i.e. the conditional probability that the region covers the parameter) is less than the unconditional confidence level uniformly in the parameters. If confidence regions are interpreted as summarizing the knowledge of a parameter value obtained from an experiment, such behavior has been considered undesirable, particularly when the conditioning subset is in some sense irrelevant to the parameter of interest. ([2], [3], [4, Chap. IV], [9].) In many of these references, the issue is discussed in terms of an associated test of significance.) Buehler [1] has formalized this behavior and studied numerous examples. Tukey [9] has given a somewhat different formalization and obtained a number of results as part of a more complex framework for statistical inference.

In this paper, a class of conditional properties is defined that includes the Buehler and Tukey definitions. Sufficient conditions for a confidence procedure to possess various properties are obtained. The main result is that if a level $\alpha$ confidence procedure yields, for all samples, posterior probability $\alpha$ for some prior probability distribution on the parameter space, then there are no subsets of the sample space, with respect to which the conditional confidence is uniformly less (or greater) than $\alpha$. A much more widely applicable, but slightly weaker, result is obtained if a sequence of prior distributions is used. The results apply to most of the classical confidence problems including discrete distribution problems and nuisance parameter problems as the Behrens-Fisher problem.

Confidence procedures for which no conditional confidence can be uniformly less (or greater) than and bounded away from the nominal level include the usual $t$, $\chi^2$, $F$, Pitman conditional location and scale, and Behrens-Fisher procedures. The "uniformly less" conclusion applies to the one-sided binomial and Poisson procedures.

Definitions and terminology are given in Section two, results are stated in Section three and proved in Sections five and six, and examples are given in Section four.

**2. Notation and definitions.** Let $Z$ be a sample space, $\Omega$ a parameter space, and $Y = Z \times \Omega$ their Cartesian product. For any set $C$ in $Y$, let $C_{z\cdot} = \{\omega: (z, \omega) \varepsilon C\}$ and $C_{\cdot\omega} = \{z: (z, \omega) \varepsilon C\}$ denote the cross section sets. Let $(Z, \mathcal{A}, \mu)$ and $(\Omega, \mathcal{B}, \lambda)$ be measure spaces with $\sigma$-finite measures $\mu$, $\lambda$. Let

Received December 29, 1958.

[1] Research carried out at the Statistical Research Center, University of Chicago, under the partial sponsorship of the Statistics Branch, Office of Naval Research. I wish to thank R. J. Buehler and F. J. Anscombe for helpful criticisms and suggestions.

$p$ be a measurable function on $Z \times \Omega$, such that for each $\omega \varepsilon \Omega$, $p_\omega(\cdot) = p(\cdot \mid \omega)$ is a probability density on $Z$ relative to $\mu$. The function $p$ is called by Tukey [9], a *specification*. The specification $p$ will be fixed throughout this paper, though other related specifications will be used. Denote by $P_\omega$ and $E_\omega$ respectively the probability measure and expectation determined by $p_\omega$ on $Z$.

A density function $\xi$ on $\Omega$ relative to the measure $\lambda$ will be called a *prior density*. More generally, if $\zeta$ is any nonnegative function on $\Omega$, not identically zero, $\zeta$ will be called a *prior quasi-density*. A prior quasi-density $\zeta$ will be said to be *admissible* with respect to the specification $p$, if

$$h_\zeta(z) \equiv \int_\Omega \zeta(\omega) p(z \mid \omega)\, d\lambda(\omega) < \infty$$

for all $z \varepsilon Z$ except for a set of $\mu$ measure zero. A prior quasi-density $\zeta$ will be said to be *admissible except on the set $A$ with respect to the specification* $p$ where $A$ is any subset of $Z$, if $h_\zeta(z) < \infty$ for all $z \varepsilon Z - A$ except for a set of $\mu$ measure zero. Every prior quasi-density for which a constant multiple is a prior density, is admissible with respect to any specification.

For every prior quasi-density $\zeta$ and every $z \varepsilon Z$ for which $0 < h_\zeta(z) < \infty$, there is defined a density $g_\zeta(\cdot \mid z)$ on $\Omega$ relative to $\lambda$:

$$g_\zeta(\omega \mid z) = \frac{p(z \mid \omega)\zeta(\omega)}{h_\zeta(z)}.$$

(That $g_\zeta$ is undefined for some $z$ will not matter.) If $\zeta$ or some constant multiple of $\zeta$ is a prior density, $g_\zeta(\cdot \mid z)$ is the *posterior density* given by Bayes theorem. If not, $g_\zeta(\cdot \mid z)$ is still a probability density on $\Omega$, but will be called here a *weak posterior density*. (Some useful simple properties of weak posterior densities are set forth in Section five.) If $\xi$ is a prior density, then $h_\xi(\cdot)$ is a (marginal) density on $Z$ relative to $\mu$.

A *confidence procedure* is a measurable set $C$ in the product space $Z \times \Omega$ with the interpretative rule that to each $z$, the confidence set $C_{z\cdot} = \{\omega: (z, \omega) \varepsilon C\}$ in $\Omega$ is assigned. Tukey calls $C$ an *event*. No restrictions concerning confidence level will be placed in the definition of a confidence procedure.

A confidence procedure $C$ is said to be *level $\alpha$ Bayes against $\xi$ with respect to the specification $p$*—written $C$ is $B(\alpha, \xi, p)$—if some constant multiple of $\xi$ is a prior density on $\Omega$ and if, for each $z \varepsilon Z$ for which $g_\xi(\cdot \mid z)$ is defined, the set $C_{z\cdot}$ has probability $\alpha$ under the posterior density $g_\xi(\cdot \mid z)$.

A confidence procedure $C$ is said to be *level $\alpha$ weak Bayes against $\zeta$ with respect to the specification $p$*—written $C$ is $B^*(\alpha, \zeta, p)$—if $\zeta$ is an admissible prior quasi-density on $\Omega$ and no multiple of $\zeta$ is a prior density, and if, for each $z \varepsilon Z$ for which $g_\zeta(\cdot \mid z)$ is defined, the set $C_{z\cdot}$ has probability $\alpha$ under the weak posterior density $g_\zeta(\cdot \mid z)$.

A confidence procedure $C$ is said to be *lower level $\alpha$ weak Bayes against $\zeta$ with respect to the specification* p—written $C$ is $B^{**}(\alpha, \zeta, p)$—if $\zeta$ is a prior quasi-density on $\Omega$ admissible except for a set $A$ (which may be empty), such that

$C_{z\cdot} = \Omega$ for all $z \varepsilon A$ and $C_{z\cdot}$ has a probability of at least $\alpha$ under the weak posterior density $g_{\mathfrak{r}}(\cdot \mid z)$ for all $z$ for which $g$ is defined.

Following Tukey [9], define a *selection* as a function $k$ mapping $Z$ into the unit interval such that $E_\omega(k) > 0$ for all $\omega \varepsilon \Omega$.[2] Let

$$p^{(k)}(z \mid \omega) = \frac{p(z \mid \omega)k(z)}{E_\omega(k)}.$$

$p^{(k)}$ is a specification and will be called the *selected (by $k$) specification*. Denote by $P_\omega^{(k)}$, $E_\omega^{(k)}$, $g_{\mathfrak{r}}^{(k)}$ the functions for the selected specifications corresponding to $P_\omega$, $E_\omega$, $g_{\mathfrak{r}}$.

Selection has the interpretation that in any conceptual infinite sequence of observation and parameter pairs, $\{(z_n, \omega_n); n = 1, 2, \cdots\}$, a new sequence is obtained as the subsequence in which the pair $(z_n, \omega_n)$ is retained according to the outcome of a chance process with retention probability $k(z_n)$. The process is assumed independent for each pair. If $k(z)$ takes only the values 0 and 1 (pure selection), the selection is according as $z_n$ does or does not belong to the set $D = \{z: k(z) = 1\}$ and the selected specification consists of the family of densities $p_\omega(\cdot)$ truncated to the sample subspace $D$.

Define, now, a number of performance properties of a confidence procedure $C$.

1. $C$ has property $c(\alpha)$ called *exact confidence* $\alpha$ if for all $\omega \varepsilon \Omega$, $P(C_{\cdot\omega}) = \alpha$.
2. $C$ has property $\underline{c}(\alpha)$ called *lower confidence* $\alpha$ if

$$\inf_{\omega \varepsilon \Omega} P_\omega(C_{\cdot\omega}) = \alpha.$$

3. $C$ has property $\bar{c}(\alpha)$ called *upper confidence* $\alpha$ if

$$\sup_{\omega \varepsilon \Omega} P_\omega(C_{\cdot\omega}) = \alpha.$$

4. $C$ has *advance probability* $\alpha$ if it has exact confidence $\alpha$, and if, for any selection $k$ for which $P_\omega^{(k)}(C_{\cdot\omega}) = q$ for all $\omega \varepsilon \Omega$, $q = \alpha$.

5. $C$ has *strong advance probability* $\alpha$ if it has advance probability $\alpha$, and if, for any selections $k_1$, $k_2$ for which

$$P_\omega^{(k_1)}(C_{\cdot\omega}) \leqq P_\omega^{(k_2)}(C_{\cdot\omega})$$

for all $\omega \varepsilon \Omega$, equality holds for all $\omega \varepsilon \Omega$.

6. $C$ has property $S_0(\alpha)$ if, for every selection $k$,

$$\alpha \leqq \sup_{\omega \varepsilon \Omega} P_\omega^{(k)}(C_{\cdot\omega}).$$

7. $C$ has property $S_1(\alpha)$ if, for every selection $k$,

$$\inf_{\omega \varepsilon \Omega} P_\omega^{(k)}(C_{\cdot\omega}) \leqq \alpha \leqq \sup_{\omega \varepsilon \Omega} P_\omega^{(k)}(C_{\cdot\omega}).$$

[2] The restriction on positivity seems possibly too strong, except when the positive domain of $p(\cdot \mid \omega)$ is the same for all $\omega$.

8. $C$ has property $S_2(\alpha)$ if, for every selection $k$, there exist parameter values $\omega_1$, $\omega_2$, such that

$$P^{(k)}_{\omega_1}(C_{\cdot\omega_1}) \leqq \alpha \leqq P^{(k)}_{\omega_2}(C_{\cdot\omega_2}).$$

9. $C$ has property $S_3(\alpha)$ if, for every selection $k$ for which $P^{(k)}_{\omega}(C_{\cdot\omega}) \leqq \alpha$ for all $\omega$ or else $P^{(k)}_{\omega}(C_{\cdot\omega}) \geqq \alpha$ for all $\omega$, equality holds for all $\omega$.

10. $C$ has property $S_4(\alpha)$ if it has property $S_3(\alpha)$ and if, for every pair of selections $k_1$, $k_2$ for which

$$P^{(k_1)}_{\omega}(C_{\cdot\omega}) \leqq P^{(k_2)}_{\omega}(C_{\cdot\omega})$$

for all $\omega$, equality holds for all $\omega$.

The properties have evident interrelations of which the most important are
strong advance probability $\alpha \Rightarrow$ advance probability $\alpha \Rightarrow c(\alpha)$.
$S_4(\alpha) \Rightarrow S_3(\alpha) \Rightarrow S_2(\alpha) \Rightarrow S_1(\alpha) \Rightarrow S_0(\alpha)$.
If $C$ has property $c(\alpha)$, then
strong advance probability $\alpha \Leftrightarrow S_4(\alpha) \Rightarrow \cdots \Rightarrow S_1(\alpha) \Rightarrow$ advance probability $\alpha$.

The ordinary term *confidence coefficient* $\alpha$ usually means exact confidence $\alpha$, or sometimes lower confidence $\alpha$. Tukey [9] introduced the sequence frequency (equivalent to exact confidence), advance probability, and strong advance probability to describe successively stronger properties of a confidence procedure in retaining "level $\alpha$" under selections. The properties $S_1(\alpha)$ and $S_2(\alpha)$ for pure selections have been defined and studied by Buehler [1]. He names the selections violating the defining condition rather than the property. The principal reason for introducing the sequence $\{S_i(\alpha)\}$ is to permit differentiation of behavior of common confidence procedures. The unsymmetric property $S_0(\alpha)$ seems of interest in much the same "conservative" way that lower confidence $\underline{c}(\alpha)$ is of interest.

Buehler's examples, combined with the examples and results of this paper, seem to indicate the need for properties intermediate to advance probability and strong advance probability, and even suggest that strong advance probability may be so strong and rare as to be of little value.

**3. Principal results.**

THEOREM 1: *Let $C$ be a confidence procedure which is level $\alpha$ Bayes against $\xi$ with respect to the specification $p$ for some $\xi$. Then $C$ is $S_2(\alpha)$. If in addition, $\xi$ is positive on $\Omega$, $C$ is $S_3(\alpha)$.*

COROLLARY 1: *A confidence procedure $C$ which has lower (or upper) confidence $\alpha$, but not exact confidence $\alpha$, is never level $\alpha$ Bayes against any $\xi$ positive on $\Omega$.*

THEOREM 2: *Let $C$ be a confidence procedure which is level $\alpha$ weak Bayes against $\zeta$ with respect to the specification $p$ for some $\zeta$. Then $C$ is $S_1(\alpha)$.*

COROLLARY 2: *If $C$ has exact confidence $\alpha$ and is level $\alpha$ weak Bayes against $\zeta$, then $C$ has advance probability $\alpha$.*

COROLLARY 3: *If a fiducial distribution for $\omega$ for the sample point $z$ has density*

*$f(\cdot \mid z)$ which is a weak posterior density $g_\zeta(\cdot \mid z)$ with respect to some admissible prior quasi-density for all $z$, then any confidence procedure giving fiducial probability $\alpha$ for every $z$ has the property $S_1(\alpha)$.*

This specifically includes results of "integrating out" nuisance parameters. Such procedures will not, in general, have any of the confidence level properties: $c(\alpha)$, $\underline{c}(\alpha)$ or $\bar{c}(\alpha)$. The result is not at all dependent on the problems of construction or meaning of fiducial distributions and fiducial probability.

The results in examples (a), (b) of Section four could have been obtained using results of Fisher and of Jeffreys ([5], [6]) together with Corollary 3, but it seems preferable to derive directly the facts necessary to apply Theorem 2.

Confidence procedures for functions of $\omega$ with nuisance parameters are easily handled directly by Theorems 1 and 2 and Corollary 2. For a more explicit treatment in an important special case, suppose $\omega = (\theta, \phi)$ with $\Omega = \Theta \times \Phi$. A confidence procedure $C$ will be called a *confidence procedure for $\theta$* if $C$ is a cylinder set with base $C^*$ in $Z \times \Theta$. Assume that the measure $\lambda$ on $\Omega$ is a product measure $\lambda_1 \times \lambda_2$ of measures on $\Theta$ and $\Phi$.

COROLLARY 4: *If a confidence procedure $C$ for $\theta$ with base $C^*$ in $Z$ has the property that*

$$C^*_z = \{\theta: (z, \theta) \varepsilon C^*\}$$

*has, for each $z$, probability $\alpha$ under the marginal distribution on $\Theta$ of a weak posterior density, then $C$ is $S_1(\alpha)$.*

THEOREM 3: *Let C be a confidence procedure which is lower level $\alpha$ weak Bayes against $\zeta$ with respect to the specification $p$ for some $\zeta$. Then $C$ is $S_0(\alpha)$.*

Proofs are given in section six.

**4. Examples.** In examples (a), (b) and (c), $Z$ is a Euclidean space with $\mu$ Lebesgue measure. In examples (d) and (e), $Z$ is the nonnegative integers with counting measure. In all examples, $\Omega$ is a Euclidean space (or obvious subspace) with $\lambda$ Lebesgue measure.

(a) *Normal.* Let $Z$ be $n$-dimensional Euclidean space with coordinates independently and identically distributed as $N(\theta, \sigma^2)$. Let $\omega = (\theta, \sigma)$, $\bar{z} = \sum z_i/n$, $S = \sum (z_i - \bar{z})^2$.

(i) $\sigma$ known. With admissible prior quasi-density $\zeta(\theta) \equiv 1$, the weak posterior density for $\theta$ is

$$g_\zeta(\theta \mid z) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n(\theta - \bar{z})^2}{2\sigma^2}}.$$

Hence any confidence procedure with confidence sets of the form

$$\{\theta: \theta - \bar{z} \varepsilon A_1\}$$

with $A_1$ a set on the real line with probability $\alpha$ under the distribution $N(0, \sigma^2/n)$ will have exact confidence $\alpha$, be weak Bayes and $S_1(\alpha)$ and have advance probability $\alpha$.

Not all such procedures are $S_2(\alpha)$. For let the set $A_1$ be any half-infinite interval, say $(-\infty, a)$. Let $k$ be a pure selection retaining the point $z$ if $\bar{z} < b$. The conditional confidence level

$$P_\theta\{z: \theta - \bar{z} < a \mid \bar{z} < b\} < \alpha$$

for all $\theta$ and $b$. The complementary selection gives conditional confidence greater than $\alpha$ for all $\theta$. $S_1(\alpha)$ guarantees that the conditional confidence is not uniformly below (or above) and bounded away from $\alpha$. I do not know if confidence procedures with $A_1$ a finite interval must have the property $S_2(\alpha)$.

(ii) $\sigma$ unknown, $n \geqq 2$. With admissible prior quasi-density $\zeta(\theta, \sigma) = 1/\sigma$, the weak posterior density is

$$g_\zeta(\theta, \sigma \mid z) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n(\theta-\bar{z})^2}{2\sigma^2}} \cdot \frac{1}{\sigma^2\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{S}{2\sigma^2}\right)^{\frac{n-1}{2}} e^{-\frac{S}{2\sigma^2}}.$$

This can be best described as follows: $(\theta, \sigma)$ given $z$ is distributed such that $S/\sigma^2$ is distributed as chi-square on $n - 1$ degrees of freedom and, conditional on $\sigma$, $\theta$ is $N(\bar{z}, \sigma^2/n)$. The marginal distribution of $\theta$ given $z$ is such that $(\theta - \bar{z})\sqrt{n(n-1)/S}$ is distributed as Student's $t$ on $n - 1$ degrees of freedom.

Any confidence procedure for $\theta$ with confidence sets of the form

$$\{\theta: (\theta - \bar{z})\sqrt{n(n-1)/S} \, \varepsilon \, A_2\}$$

with $A_2$ a set on the real line with probability $\alpha$ under the $t_{n-1}$ distribution will have exact confidence $\alpha$, and, using Corollary 4, will be weak Bayes level $\alpha$ with respect to $\zeta$. The procedure is then $S_1(\alpha)$ and has advance probability $\alpha$.

Buehler [1] has noted that *no* such procedure is $S_2(\alpha)$, a pure selection according as $S \geqq c$ giving conditional confidence uniformly less than $\alpha$.

Any confidence procedure for $\sigma$ with confidence sets of the form

$$\{\sigma: S/\sigma^2 \, \varepsilon \, A_3\}$$

with $A_3$ a set on the positive real line with probability $\alpha$ under the $\chi^2_{n-1}$ distribution, will have exact confidence $\alpha$, be weak Bayes, and hence $S_1(\alpha)$ and have advance probability $\alpha$.

These results are all special cases of example (c) on location and scale parameters.

(b) *The Behrens-Fisher problem.* Let $Z$ be $n_1 + n_2$ dimensional Euclidean space, with coordinates independently distributed, the first $n_1$ identically as $N(\theta_1, \sigma_1^2)$, the last $n_2$ identically as $N(\theta_2, \sigma_2^2)$. Let $\omega = (\theta_1, \sigma_1, \theta_2, \sigma_2)$ and let $\bar{z}_1, \bar{z}_2, S_1, S_2$ be the means and sums of squares of deviations of the two sets of coordinates. Assume $n_1 \geqq 2$, $n_2 \geqq 2$.

With the admissible prior quasi-density $\zeta(\omega) = 1/\sigma_1\sigma_2$, the weak posterior distributions of $(\theta, \sigma_1)$ and $(\theta_2, \sigma_2)$ are independent and as obtained in example $(a_{ii})$ with appropriate $(n_i, \bar{z}_i, S_i)$.

The usual confidence procedures for $\sigma_2/\sigma_1$ with confidence sets of the form

$$\left\{ \sigma_2/\sigma_1 : \frac{S_1}{(n_1 - 1)\sigma_1^2} \cdot \frac{(n_2 - 1)\sigma_2^2}{S_2} \, \varepsilon \, A_1 \right\}$$

with $A_1$ having probability $\alpha$ under the $F_{n_1-1, n_2-1}$ distribution will have exact confidence $\alpha$, be weak Bayes and $S_1(\alpha)$ and have advance probability $\alpha$.

The marginal weak posterior distribution for $\theta_1 - \theta_2$ is easily found to be such that

$$\frac{(\theta_1 - \theta_2) - (\bar{z}_1 - \bar{z}_2)}{\sqrt{a_1 + a_2}}$$

is distributed as the linear combination of independent Student's variates: $t_{n_1-1} \sin \theta - t_{n_2-1} \cos \theta$, where $a_i = S_i/[n_i(n_i - 1)]$ and $\theta = \tan^{-1}[a_1/a_2]^{\frac{1}{2}}$. This distribution can usefully be called the Behrens-Fisher distribution with parameters $n_1 - 1$, $n_2 - 1$ and $\theta$ — written $\mathrm{BF}(n_1 - 1, n_2 - 1; \theta)$. (The usefulness of this terminology is illustrated in the paper [10] in which a more detailed related treatment of the Behrens-Fisher problem is given.)

Any confidence procedure for $\theta_1 - \theta_2$ with confidence sets of the form

$$\left\{ \theta_1 - \theta_2 : \frac{(\theta_1 - \theta_2) - (\bar{z}_1 - \bar{z}_2)}{\sqrt{a_1 + a_2}} \, \varepsilon \, A_2 \right\}$$

with $A_2$ a set having probability $\alpha$ under the distribution $\mathrm{BF}(n_1 - 1, n_2 - 1, \theta)$ will be weak Bayes level $\alpha_1$ and have the property $S_1(\alpha)$. Since the marginal weak posterior density for $\theta_1 - \theta_2$ is exactly the fiducial density for $\theta_1 - \theta_2$ under the Behrens-Fisher solution, and fiducial procedure with fiducial probability $\alpha$ has the property $S_1(\alpha)$. Such procedures are known not to have exact confidence $\alpha$, but at least for $n_1$ and $n_2$ sufficiently large to have lower confidence $\alpha$ and not be $S_2(\alpha)$. The behavior for small $n_1$ and $n_2$ is unclear. (C.f. [10].)

The Welch asymptotic procedure with asymptotically exact confidence $\alpha$ does not possess property $S_1(\alpha)$. Fisher's criticism ([3]) of this procedure amounts effectively to showing that, for $n_1 = n_2 = 7$, a pure selection with retention if $|(S_1/S_2) - 1| < \delta$ for $\delta$ small gives conditional confidence uniformly below and bounded away from $\alpha$. For $n_1$ and $n_2$ sufficiently large, asymptotic theory suffices to show that a selection with retention if

$$|[S_1 n_2(n_2 - 1)^2 / S_2 n_1(n_1 - 1)^2] - 1| < \delta$$

for $\delta$ small has a similar effect. Calculations for small and moderate values of $n_1$ and $n_2$ indicate that the effect holds fairly generally.

(c) *Location and scale parameter families.* Let $Z$ be $n$-dimensional Euclidean space, let $\omega = (\theta, \sigma)$ with $\theta$ a (real) location parameter, $\sigma$ a (positive) scale parameter for the family $p$ of distributions. Let $\epsilon = (1, \cdots, 1)$. Then

$$p_\omega(z) = \frac{1}{\sigma^n} \, q\left(\frac{z - \theta\epsilon}{\sigma}\right)$$

for a fixed density $q$.

(i) $\sigma = 1$, known. The prior quasi-density $\zeta_1(\theta) = 1$ is admissible ($\int q(z - \theta\epsilon)\, d\theta < \infty$ except on sets of measure zero) for all $q$, and the weak posterior density of $\theta$ is:

$$g_{\zeta_1}(\theta \mid z) = \frac{q(z - \theta\epsilon)}{\int_{-\infty}^{\infty} q(z - \psi\epsilon)\, d\psi}.$$

Let $C$ be a confidence procedure which is level $\alpha$ weak Bayes with respect to $\zeta_1$ and which also possesses the translation property: $(z, \theta) \varepsilon C$ if and only if, $(z + a\epsilon, \theta + a) \varepsilon C$ for all $a$, or equivalently, if and only if $z - \theta\epsilon \varepsilon C_{\cdot 0}$. The translation property guarantees that the confidence level is constant; for

$$\int_{C_{\cdot\theta}} q(z - \theta\epsilon)\, d\mu(z) = \int_{C_{\cdot 0}} q(z)\, d\mu(z)$$

and since $C$ is $S_1(\alpha)$, then it must have exact confidence $\alpha$, and have advance probability $\alpha$. Such a procedure is a conditional procedure with translation property as constructed by Pitman [7], by choosing a set with conditional probability $\alpha$ under the conditional distribution of $z - \theta\epsilon$ on each configural line determined by the differences $\{z_i - z_1,\ i = 2, \cdots, n\}$ of the sample point. Buehler [1] proved the $S_1(\alpha)$ result when $n = 1$.

(ii) $\theta = 0$, known. The prior quasi-density $\zeta_2(\sigma) = 1/\sigma$ is admissible for all $q$, and the weak posterior density of $\sigma$ is

$$g_{\zeta_2}(\sigma \mid z) = \frac{\sigma^{-(n+1)} q(z/\sigma)}{\int_{-\infty}^{\infty} \tau^{-(n+1)} q(z/\tau)\, d\tau}.$$

Again, a confidence procedure which is level $\alpha$ weak Bayes with respect to $\zeta_2$ and has the natural property under scale change, has exact confidence $\alpha$, is $S_1(\alpha)$ and has advance probability $\alpha$. It is a Pitman scale procedure, with conditional confidence $\alpha$ on each configural ray from the origin.

(iii) For $n \geqq 2$, the prior quasi-density $\zeta_3(\theta, \sigma) = 1/\sigma$ is admissible for all $q$ and the weak posterior density of $(\theta, \sigma)$ is

$$g_{\zeta_3}(\theta, \sigma \mid z) = \frac{\sigma^{-(n+1)} q\left(\dfrac{z - \theta\epsilon}{\sigma}\right)}{\int_{-\infty}^{\infty} d\psi \int_0^{\infty} d\tau \left[\tau^{-(n+1)} q\left(\dfrac{z - \psi}{\tau}\right)\right]}.$$

Let $C$ be a confidence procedure which is level $\alpha$ weak Bayes with respect to $\zeta_3$ and which has the translation scale change property that $(z, (\theta, \sigma)) \varepsilon C$ if and only if $(z - \theta\epsilon)/\sigma \varepsilon C_{\cdot(0,1)}$. Any such procedure will be $S_1(\alpha)$, have exact confidence $\alpha$ and advance probability $\alpha$. Included are confidence procedures for $\theta$ and $\sigma$ jointly and for $\theta$ or $\sigma$ separately. For the latter, it suffices to determine each $C_{z\cdot}$ according to the marginal posterior densities. Again, the procedures are just those of Pitman.

In all examples, the prior quasi-densities were those of the Haar measure on

the appropriate group of translation and/or scale changes. If $\Omega$ is any $\sigma$-compact group of transformations on $Z$, and if $\zeta(\omega)\, d\lambda(\omega)$ is Haar measure on $\Omega$, then a confidence procedure $C$ can be obtained which is weak Bayes level $\alpha$ and which satisfies the standard invariance condition under the group $\Omega$. The latter condition insures that the procedure has exact confidence not depending on $\omega$ and, from the former, the procedure is $S_1(\alpha)$, hence has exact confidence $\alpha$ and advance probability $\alpha$. To prove that the Haar measure is admissible and to show that the weak posterior density is for each fixed $\omega$ the conditional density on each orbit in $Z$ under the group $\Omega$, seems to require slightly more structure.

(d) *Binomial.* Let $p_\omega$ be the binomial $(n, \omega)$ density. The usual one-sided confidence interval $(l_1(z), 1)$ for $\omega$ with lower confidence $\alpha$ is obtained for $z > 0$ as the root of the equation

$$\sum_{i=0}^{z-1} p(i \mid l_1) = \sigma$$

or, equivalently, of the incomplete beta equation

$$\frac{1}{B(z, n-z+1)} \int_{l_1}^{1} \omega^{z-1}(1-\omega)^{n-z}\, d\omega = \alpha.$$

For $z = 0$, $l_1(0) = 0$. The weak posterior density with respect to the prior quasi-density $\zeta_1(\omega) = \omega^{-1}$ (i.e., a beta quasi-distribution with parameters $(0, 1)$) is

$$g_{\zeta_1}(\omega \mid z) = \frac{\omega^{z-1}(1-\omega)^{n-z}}{B(z, n-z+1)}.$$

Since $\zeta_1$ is admissible except for $z = 0$, for which $C_{0\cdot} = \Omega$, the confidence procedure is lower level $\alpha$ weak Bayes and is $S_0(\alpha)$. Selection by the set $\{z = 0\}$ shows that the procedure is not $S_1(\alpha)$.

Similarly, the usual confidence interval $(0, l_2(z))$ with lower confidence $\alpha$ is lower level $\alpha$ weak Bayes with respect to the prior quasi-density $\zeta_2(\omega) = (1-\omega)^{-1}$, and is $S_0(\alpha)$ but not $S_1(\alpha)$.

The usual two-sided confidence interval $(l_1(z), l_2(z))$ combining the two one-sided lower confidence $\alpha$ procedures is a much more complex procedure having lower confidence depending on $n$ and $\alpha$, but lying between $2\alpha - 1$ and $\alpha$. The nominal lower level $2\alpha - 1$ is achieved only for rare combinations of $n$ and $\alpha$, so that, *a fortiori*, the two-sided procedure is not usually $S_1(2\alpha - 1)$

(e) *Poisson*: Let $p_\omega$ be the Poisson density with mean $\omega$. As with the binomial, the usual one-sided procedure for $\omega$ with intervals of the form $(l_1(z), \infty)$ and with lower confidence $\alpha$ is lower level $\alpha$ weak Bayes against the prior quasi-density $\zeta_1(\omega) = \omega^{-1}$, and the procedure is $S_0(\alpha)$ Since $l_1(0) = 0$, it is not $S_1(\alpha)$. However, the other one-sided procedure does not suffer from the end effect and the interval $(0, l_2(z))$ determined for all $z$ from the equation

$$\sum_{z+1}^{\infty} p(i \mid l_2) = \alpha$$

or from the equation

$$\frac{1}{\Gamma(z+1)} \int_0^{t_2} \omega^z e^{-\omega} \, d\omega = \alpha$$

gives a confidence procedure with lower confidence $\alpha$ which is level $\alpha$ weak Bayes against the prior quasi-sensity $\zeta_2(\omega) = 1$ and hence has the property $S_1(\alpha)$.

**5. A property of prior quasi-densities and weak posterior densities.**

THEOREM 3: *If $\zeta$ is a prior quasi-density with $\int \zeta(\omega) \, d\lambda = \infty$, with corresponding weak posterior density $g_\zeta(\cdot \mid z)$, then there exists a sequence of prior densities $\{\xi_n\}$ with corresponding posterior densities $g_n(\cdot \mid z)$ such that for all $\omega \varepsilon \Omega$ and all $z$ for which $g_\zeta$ is defined,*

$$\lim_{n\to\infty} g_n(\omega \mid z) = g_\zeta(\omega \mid z).$$

*Further, if $\{\xi_n\}$ is any sequence of prior densities with corresponding prior densities $\{g_n(\cdot \mid z)\}$ such that there exist constants $K$ and $\{a_n ; n = 1, 2, \cdots\}$ such that for all $\omega$,*

$$\lim_{n\to\infty} a_n\xi_n(\omega) = \zeta(\omega) \tag{5.1}$$

*and*

$$a_n\xi_n(\omega) \leqq K\zeta(\omega) \tag{5.2}$$

*then*

$$\lim_{n\to\infty} g_n(\omega \mid z) = g_\zeta(\omega \mid z).$$

In the second part of the theorem, condition (5.2) is necessary in that sequences $\{\xi_n\}$ can be found that satisfy condition (5.1) but for which $\{g_n(\omega \mid z)\}$ does not converge, or which converges but not to a probability density.

The second part will be proved first. For all $z$ for which the right-hand denominator is finite and positive,

$$g_n(\omega \mid z) = \frac{a_n p(z \mid \omega)\xi_n(\omega)}{a_n \int_\Omega p(z \mid u)\xi_n(u) \, d\lambda(u)} .$$

Under conditions (5.1) and (5.2), both numerator and denominator converge respectively to the numerator and denominator of $g_\zeta(\omega \mid z)$ for all $\omega$ and for every $z$ for which $g_\zeta$ is defined.

The first part of the theorem will be proved by exhibiting a sequence $\{\xi_n\}$ satisfying conditions (5.1) and (5.2). Since $\lambda$ is $\sigma$-finite, there exists an increasing sequence of sets in $\Omega$:$\{B_n ; n = 1, \cdots\}$ such that $\lim B_n = \Omega$ and $\lambda(B_n) < \infty$.

Define

$$\xi_n(\omega) = \begin{cases} \dfrac{\min\,[n, \zeta(\omega)]}{\displaystyle\int_{B_n} \min\,[n, \zeta(u)]\, d\lambda(u)} & \omega \,\varepsilon\, B_n \\ 0 & \omega \,\varepsilon\, B_n \end{cases}$$

$\xi_n$ clearly satisfies the two conditions with $K = 1$,

$$a_n = \int_{B_n} \min\,[n, \zeta(u)]\, d\lambda(u).$$

**6. Proofs of results of section three.**

LEMMA 1: *If a confidence procedure $C$ is $B(\alpha, \xi, p)$ then for every selection $k$, $C$ is $B(\alpha, \xi_k, p^{(k)})$, with prior density*

$$\xi_k(\omega) = \frac{\xi(w)\cdot E_\omega(k)}{\displaystyle\int \xi(u)\cdot E_u(k)\, d\lambda(u)}.$$

LEMMA 2: *If a confidence procedure $C$ is $B^*(\alpha, \zeta, p)$ (or $B^{**}(\alpha, \zeta, p)$), then for every selection $k$, $C$ is $B^*(\alpha, \zeta_k, p^{(k)})$ (or $B^{**}(\alpha, \zeta_k, p^{(k)})$) with prior quasi-density*

$$\zeta_k(\omega) = \zeta(\omega)\cdot E_\omega(k).$$

By assumption in Lemma 1,

$$\frac{\displaystyle\int_{C_z.} \xi(\omega) p(z \mid \omega)\, d\lambda(\omega)}{\displaystyle\int \xi(u) p(z \mid u)\, d\lambda(u)} = \alpha.$$

But

$$\xi_k(\omega) p^k(z \mid \omega) = a(z)\xi(\omega)p(z \mid \omega)$$

so that

$$\frac{\displaystyle\int_{C_z.} \xi_k(\omega) p^{(k)}(z \mid \omega)\, d\lambda(\omega)}{\displaystyle\int_{\Omega} \xi_k(u) p^{(k)}(z \mid u)\, d\lambda(u)} = \alpha.$$

The proof of Lemma 2 is the same with the exceptional set for admissibility unchanged for the selected specfication.

Let $\chi_C$ denote the set characteristic function of the set $C$ in $Z \times \Omega$. By the usual conditional expectation interchange of order of integration, for any prior density $\xi$ and specification $p$,

$$
\begin{aligned}
\int_\Omega P_\omega(C,\omega)\xi(\omega)\,d\lambda(\omega) &= \int_{Z\times\Omega} \chi_C(z,\omega)p(z\mid\omega)\xi(\omega)\,d[(\mu\times\lambda)(z,\omega)] \\
&= \int_Z h_\xi(z)\left[\int_\Omega \chi_C(z,\omega)g_\xi(\omega\mid z)\,d\lambda(\omega)\right]d\mu(z) \qquad (6.1)\\
&= \int_Z h_\xi(z)\left[\int_{C_{z.}} g_\xi(\omega\mid z)\,d\lambda(\omega)\right]d\mu(z).
\end{aligned}
$$

Note that the lack of definition of $g_\xi$ for those $z$ for which $h_\xi(z) = 0$ is of no consequence.

If $C$ is $B(\alpha, \xi, p)$, then for any selection $k$, $C$ is $B(\alpha, \xi_k, p^{(k)})$ by Lemma 1, and applying equation (6.1) to $\xi_k$ and specification $p^{(k)}$ yields

$$\int_\Omega P_\omega^{(k)}(C_{.\omega})\xi_k(\omega)\,d\mu(z) = \alpha.$$

It follows immediately that $P_\omega^{(k)}(C_{\cdot\omega}) < \alpha(>\alpha)$ for all $\omega$ is impossible so $C$ is $S_2(\alpha)$. If $\xi$ is positive on $\Omega$, so is $\xi_k$, and $C$ is $S_3(\alpha)$ and Theorem 1 is proved.

If $C$ is $B^*(\alpha, \zeta, p)$, then for any selection $k$, $C$ is $B^*(\alpha, \zeta_k, p^{(k)})$ by Lemma 2. Let $\{\xi_n\}$ be the sequence of prior densities guaranteed by Theorem 4 and let $\{g_n(\cdot\mid z)\}$ be the corresponding posterior densities under $p^{(k)}$, converging to $g_{\zeta_k}^{(k)}(\cdot\mid z)$. Since these are probability densities on $\Omega$, it follows from Scheffé's theorem [8] that

$$\lim_{n\to\infty}\int_{C_{z.}} g_n(\omega\mid z)\,d\lambda(\omega) = \int_{C_{z.}} g_{\zeta_k}(\omega\mid z)\,d\lambda(\omega) \qquad (6.2)$$

uniformly in $z$ with the right hand side identically equal to $\alpha$ by hypothesis. Then with

$$h_n(z) = \int \xi_n(\omega)p^{(k)}(\omega\mid z)\,d\lambda(\omega),$$

$$\lim_{n\to\infty}\int_Z h_n(z)\left[\int_{C_{z.}} g_n(\omega\mid z)\,d\lambda(\omega)\right]d\mu(z) = \alpha,$$

and this, together with equation (6.1) applied to $\xi_n$ and $p^{(k)}$, yields

$$\lim_{n\to\infty}\int_\Omega P_\omega^{(k)}(C_{.\omega})\xi_n(\omega)\,d\lambda(\omega) = \alpha.$$

Hence, for no $k$ is it possible that

$$\sup_\omega P_\omega^{(k)}(C_{\cdot\omega}) < \alpha$$

or

$$\inf_\omega P_\omega^{(k)}(C_{\cdot\omega}) > \alpha$$

so that $C$ is $S_1(\alpha)$ and Theorem 2 is proved.

If $C$ is $B^{**}(\alpha, \zeta, p)$ with $\zeta$ admissible except for the set $A \subset Z$ then, using the same notation as in the preceding proof, equation (6.2) holds uniformly in $z$ for $z \not\in A$. The right hand side is now not less than $\alpha$. Then

$$\int_Z h_n(z) \left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z) - \alpha$$
$$= (1 - \alpha) \int_A h_n(z)\, d\mu(z) + \int_{Z-A} h_n(z) \left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega) - \alpha\right] d\mu(z)$$

so that

$$\liminf_{n\to\infty} \int_Z h_n(z) \left[\int_{C_{z.}} g_n(\omega \mid z)\, d\lambda(\omega)\right] d\mu(z) \geqq \alpha$$

and hence

$$\liminf_{n\to\infty} \int_\Omega P_\omega^{(k)}(C_{.\omega})\xi_n(\omega)\, d\lambda(\omega) \geqq \alpha.$$

Then for no $k$ is it possible that

$$\sup_\omega P^{(k)}(C_{.\omega}) < \alpha$$

and $C$ is $S_0(\alpha)$ and Theorem 3 is proved.

Corollary 1 follows immediately from Theorem 1, and Corollaries 2 and 3 from Theorem 2. Corollary 4 follows from Theorem 2, by noting that $C_{z.}$ is a cylinder set with base $C_{z.}^*$ in $Z \times \Theta$.

## REFERENCES

[1] R. J. Buehler, "Some validity criteria for statistical inferences," *Ann. Math. Stat.*, Vol. 30 (1959) pp. 845–863.
[2] D. R. Cox, "Some problems connected with statistical inference," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 357–372.
[3] R. A. Fisher, "On a test of significance in Pearson's Biometrika Table (No. 11)," *J. Roy. Stat. Soc., Series B*, Vol. 18 (1956), pp. 56–60.
[4] R. A. Fisher, *Statistical Methods and Scientific Inference*, Edinburgh and London, Oliver and Boyd, 1956.
[5] H. Jeffreys, *The Theory of Probability, Second Edition*, Oxford, Oxford University Press, 1948, pp. 120–122.
[6] H. Jeffreys, "Note on the Behrens-Fisher formula," *Ann. Eugenics*, Vol. 10 (1940), pp. 48–51.
[7] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1938), pp. 391–421.
[8] H. Scheffé, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.
[9] J. W. Tukey, "Fiducial inference," 1958 Wald Lectures, Cambridge, Massachusetts, meeting of the Institute of Mathematical Statistics, August 1958.
[10] D. L. Wallace, "The Behrens-Fisher problem—a critical review and a subjective approach," in preparation.

# AN EXAMPLE OF WIDE DISCREPANCY BETWEEN FIDUCIAL AND CONFIDENCE INTERVALS

BY CHARLES STEIN

*Stanford University*[1]

**1. Introduction.** Fisher [1], [2] has emphasized that when he chooses a set in the parameter space on the basis of certain observations and attributes to it a certain fiducial probability $\alpha$, he does not intend that, for fixed values of the parameter the probability that this random set contains the parameter point should be $\alpha$. Examples of this distinction for the Behrens-Fisher problem have been given by Fisher [1], [2] and Neyman [3], [4]. In these cases the numerical differences are not extremely large. In order to bring out more clearly the contrast between fiducial probability and confidence sets I shall give, for each $\alpha$ and $\epsilon$ in the interval $(0, 1)$, an example where a fiducial interval for a parameter with fiducial probability equal to $\alpha$ has probability less than $\epsilon$ of covering the true parameter for a large range of parameter values. This means that although a large fiducial probability is claimed, it is practically certain that the interval will not cover the true parameter value. Of course this cannot happen when the fiducial sets are obtained by Pitman's methods [6], [7].

**2. The example.** Let $X_1, \cdots, X_n$ be independently normally distributed real random variables with unknown means $\xi_1, \cdots, \xi_n$ and variance 1. Suppose we are interested in fiducial or confidence sets for $\sum \xi_i^2$ of the form

$$[f(X_1, \cdots, X_n), \infty].$$

We consider the one-sided case only in order to avoid irrelevant computational details. The fiducial distribution of $\xi_1, \cdots, \xi_n$ is that they are independently normally distributed with means $X_1, \cdots, X_n$ and variance 1 (see Fisher [1], p. 132, where the case $n = 2$ is given, but see also Tukey [5] for a different fiducial distribution). Thus the fiducial distribution of $\sum_1^n \xi_i^2$ is a non-central $\chi^2$ distribution with $n$ degrees of freedom and non-centrality parameter $\sum_1^n X_i^2$. On the basis of this we determine a fiducial interval

$$[\Phi_{\alpha,n}(\textstyle\sum X_i^2), \infty) \tag{1}$$

with fiducial probability $\alpha$ for the unknown parameter $\sum \xi_i^2$. Here $\Phi_{\alpha,n}(\sum X_i^2)$ is the value which will be exceeded with probability $\alpha$ by a non-central $\chi^2$ variate with $n$ degrees of freedom and non-centrality parameter $\sum X_i^2$. But this non-central $\chi^2$ distribution is, for large $n$, approximately a normal distribution with mean $n + \sum X_i^2$ and variance $2n + 4\sum X_i^2$, the approximation being uniform

Received January 12, 1959.

[1] This work was done in part at the University of California, Berkeley and in part at Stanford University, with sponsorship of the Office of Naval Research.

in $\sum X_i^2$. Thus for sufficiently large $n$

$$\Phi_{\alpha,n}(\sum X_i^2) > n + \sum X_i^2 - t'_\alpha \sqrt{2n + 4 \sum X_i^2}, \tag{2}$$

for all values of $\sum X_i^2$ where $t'_\alpha$ is a fixed number independent of $n$ satisfying

$$t'_\alpha > t_\alpha \tag{3}$$

where

$$1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{t_\alpha}^{\infty} e^{-\frac{1}{2}u^2}\, du. \tag{4}$$

Thus for fixed $\xi_1, \cdots, \xi_n$ the probability that the fiducial interval will cover the true value of $\sum \xi_i^2$ is

$$\begin{aligned} & P_{\xi_1 \cdots \xi_n}\{\sum \xi_i^2 \geqq \Phi_{\alpha,n}(\sum X_i^2)\} \\ (5) \qquad & \leqq P_{\xi_1 \cdots \xi_n}\{\sum \xi_i^2 \geqq n + \sum X_i^2 - t'_\alpha \sqrt{2n + 4\sum X_i^2}\} \\ & = P_{\xi_1 \cdots \xi_n}\{\sum X_i^2 \leqq \sum \xi_i^2 - n + t'_\alpha \sqrt{2n + 4 \sum X_i^2}\} \end{aligned}$$

for sufficiently large $n$. Now let $n \to \infty$ with

$$\lim_{n \to \infty} \frac{1}{n^2} \sum_1^n \xi_i^2 = 0. \tag{6}$$

From Chebyshev's inequality it follows that, for any $\epsilon > 0$, we have, for sufficiently large $n$,

$$P\{2n + 4 \sum X_i^2 > \epsilon n^2\} < \epsilon. \tag{7}$$

Thus

$$\begin{aligned} (8) \qquad & P_{\xi_1 \cdots \xi_n}\{\sum X_i^2 \leqq \sum \xi_i^2 - n + t'_\alpha \sqrt{2n + 4 \sum X_i^2}\} \\ & \qquad \leqq P_{\xi_1 \cdots \xi_n}\{\sum X_i^2 \leqq \sum \xi_i^2\} + \epsilon, \end{aligned}$$

Again applying Chebyshev's inequality, or the limiting distribution of $\sum X_i^2$. it follows from (6) and (8) that

$$\lim_{n \to \infty} P_{\xi_1 \cdots \xi_n}\{\sum \xi_i^2 \geqq \Phi_{\alpha,n}(\sum X_i^2)\} = 0. \tag{9}$$

Let us compare these results with the natural confidence sets. Since $\sum X_i^2$ has a non-central $\chi^2$ distribution with $n$ degrees of freedom and non-centrality parameter $\sum \xi_i^2$, the confidence sets of the desired form are

$$[\Phi_{1-\alpha,n}^{-1}(\sum X_i^2), \infty), \tag{10}$$

or, approximately for large $n$,

$$\sum X_i^2 \leqq \sum \xi_i^2 + n + t_\alpha \sqrt{2n + 4 \sum \xi_i^2} \tag{11}$$

(which must be inverted to obtain an explicit lower confidence bound for $\sum \xi_i^2$) as compared with the fiducial interval

$$[\Phi_{\alpha,n}(\textstyle\sum X_i^2),\ \infty) \tag{12}$$

which is approximately, for large $n$

$$\sum \xi_i^2 \geqq \sum X_i^2 + n - t_\alpha \sqrt{2n + 4\sum X_i^2}\,. \tag{13}$$

However a different argument leads to a more reasonable fiducial distribution. Because of the rotational symmetry of the problem, it seems reasonable to base our procedure only on $Z = \sum X_i^2$, ignoring the individual observations. Then the fiducial argument leads to the intervals based on (10), i.e. confidence intervals.

At first I intended to write this paper without extended comments, letting the example speak for itself. However some remarks of the editor and referees and the fact that I have since read the discussion of fiducial inference in Chapter 6 of Quenouille [8] lead me to believe that some discussion may be useful. Two questions may be asked in connection with the above example. Is the argument used a fiducial argument as this is understood by the advocates of fiducial inference, and is the resulting fiducial distribution of $\sum\xi_i^2$ absurd? The fiducial argument has two parts. First the joint fiducial distribution of $\xi_1 \ldots \xi_n$ is given on the authority of [1]. Then the distribution of $\sum\xi_i^2$ is calculated from this joint distribution, the joint fiducial distribution being treated as an ordinary probability distribution. The first step seems to be in agreement with the practice of the advocates of fiducial inference. For example, Quenouille on p. 139 of [8] argues against the different fiducial distribution given by Tukey [5]. Anyone who argues that the second step is not justified seems to be saying that fiducial distributions cannot be treated as ordinary probability distributions. In [8] on pp. 114–119, Quenouille imposes restrictions on the way some fiducial distributions can be used, but (at least to me) it is not clear whether these restrictions are meant to apply to cases as simple as the one discussed in this paper, nor is it clear whether my derivation meets his requirements if they are applicable to the present case.

Finally it may be contended that the fiducial interval (1) is the correct one and should be used. Because of the conflict with the argument immediately below (13), I do not think many people will take this attitude. Apart from this, there is an important question of principle here. If $n$ is large and $\sum\xi_i^2$ is small compared with $n^2$ (which is commonly the case if the $\xi_i$ are coordinates of a high order interaction), then the probability that the fiducial interval (1) will cover the true value of $\sum\xi_i^2$ has been shown to be small if $\alpha$ is moderate. This has the practical interpretation that, when the fiducial interval (1) is applied in such situations, it will not cover the true value in the vast majority of cases that actually arise. For this reason I cannot understand the contention that the probability of covering a fixed parameter point is irrelevant to inferences of this type.

## REFERENCES

[1] Ronald A. Fisher, *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh, 1956.

[2] Ronald A. Fisher, "On a point raised by M. S. Bartlett on fiducial probability," *Ann. Eugen.*, Vol. 7(1937), pp. 370–375.

[3] Jerzy Neyman, "Fiducial argument and the theory of confidence intervals," *Biometrika*, Vol. 32 (1941–42), pp. 128–150.

[4] Jerzy Neyman, *Lectures and Conferences on Mathematical Statistics and Probability*, Graduate School, U.S. Dept. Agric., Washington, Second Edition, 1952.

[5] J. W. Tukey, "Some examples with fiducial relevance," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 687–695.

[6] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939) pp. 391–421.

[7] E. J. G. Pitman, "Statistics and science," *J. Amer. Stat. Assn.*, Vol. 52 (1957), pp. 322–330.

[8] M. H. Quenouille, *The Fundamentals of Statistical Reasoning*, Charles Griffin and Co., Ltd., London, 1958.

# OPTIMUM INVARIANT TESTS[1]

By E. L. Lehmann

*University of California, Berkeley*

**Summary.** The standard (likelihood ratio) test of the general linear hypothesis has been shown to possess numerous different optimum properties. A brief survey of these was included in a recent paper by Kiefer [2]. In the present note it is shown that all of these, and in fact a wide class of optimum properties of which the above are special cases, are consequences of the fact that the test is uniformly most powerful invariant.

**1. Order relations among tests.** Let $X$ be a random variable with possible distributions $\mathcal{P} = \{P_\theta, \theta \, \varepsilon \, \Omega\}$ and consider the hypothesis $H: \theta \, \varepsilon \, \omega$ where $\omega$ is a subset of $\Omega$. Suppose that the problem of testing $H$ against the alternatives $K: \theta \, \varepsilon \, \Omega - \omega$ remains invariant under a group $G$ of transformations of the sample space. Let $\mathfrak{I}$ be a class of tests $\varphi$ of $H$, for example the class of all level $\alpha$ tests or of all unbiased level $\alpha$ tests, which is invariant under $G$ in the sense that $\varphi \in \mathfrak{I}$ implies $\varphi g \, \varepsilon \, \mathfrak{I}$ for all $g \, \varepsilon \, G$. Here $\varphi g$ denotes the critical function defined by

$$\varphi g(x) = \varphi(gx).$$

Suppose that a relation $\lesssim$ has been defined among the tests of $\mathfrak{I}$ such that every pair $\varphi, \varphi' \, \varepsilon \, \mathfrak{I}$ satisfies either $\varphi \lesssim \varphi'$ or $\varphi' \lesssim \varphi$. When both of these relations hold, we write $\varphi \approx \varphi'$. Let the (weak) ordering $\lesssim$ satisfy the following conditions:

(i) If $\varphi'$ is uniformly at least as powerful as $\varphi$, then $\varphi \lesssim \varphi'$.

(ii) If $\varphi_\gamma, \gamma \, \varepsilon \, \Gamma$ is any family of tests belonging to $\mathfrak{I}$ and $\nu$ any probability measure over the label space $\Gamma$, then $\varphi \lesssim \varphi_\gamma$ for all $\gamma \, \varepsilon \, \Gamma$ implies $\varphi \lesssim \int \varphi_\gamma \, d\nu(\gamma)$.

(iii) If $\varphi_0 \lesssim \varphi_n$ for $n = 1, 2, \cdots$ and if $\varphi$ is a critical function such that the power-functions $\beta_{\varphi_n}(\theta) \to \beta_\varphi(\theta)$ for all $\theta \, \varepsilon \, \Omega$ as $n \to \infty$, then $\varphi_0 \lesssim \varphi$.

(iv) If $\varphi \lesssim \varphi'$ then $\varphi g \lesssim \varphi' g$ for all $g \, \varepsilon \, G$.

A test $\varphi_0 \, \varepsilon \, \mathfrak{I}$ will be called *optimum* within $\mathfrak{I}$ according to this ordering if $\varphi \lesssim \varphi_0$ for all $\varphi \, \varepsilon \, \mathfrak{I}$.

The following are some examples of such orderings, which have been considered in the literature. Throughout, $\beta_\varphi$ denotes the power function of $\varphi$.

*Example* 1. Let $a(\theta) \geqq 0$ and $b(\theta)$ be functions which are invariant under the transformations $\bar{G}$ induced by $G$ in the parameter space, and let $\varphi \lesssim \varphi'$ if

$$\inf_{\Omega-\omega} [a(\theta)\beta_\varphi(\theta) + b(\theta)] \leqq \inf_{\Omega-\omega} [a(\theta)\beta_{\varphi'}(\theta) + b(\theta)].$$

Then conditions (i) to (iv) are clearly satisfied. A particular case is obtained by putting $b(\theta) = 0$; $a(\theta) = 1$ if $\theta \, \varepsilon \, \omega'$ and $a(\theta) = 0$ otherwise, where $\omega'$ is

Received September 15, 1958.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

an invariant subset of $\Omega - \omega$. Then $\varphi \lesssim \varphi'$ if

$$\inf_{\omega'} \beta_\varphi(\theta) \leqq \inf_{\omega'} \beta_{\varphi'}(\theta).$$

A test is optimum according to this ordering if it maximizes the minimum power over $\omega'$.

*Example* 2. Let the tests be ordered according to $-s(\varphi)$ where $s(\varphi)$ is the stringency of $\varphi$ defined by

$$s(\varphi) = \sup_{\Omega-\omega} [\beta^*(\theta) - \beta_\varphi(\theta)]$$

with $\beta^*$ denoting the envelope power function. Then $\varphi \lesssim \varphi'$ if $s(\varphi) \geqq s(\varphi')$ and the four conditions are again easily verified.

*Example* 3. Let $\omega'$ be an invariant subset of $\Omega - \omega$ and suppose that there exists a probability distribution $\lambda$ over $\omega'$ which is invariant under the group $\bar{G}$ induced by $G$ in the parameters face. Then the relation $\varphi \lesssim \varphi'$ if

$$\int_{\omega'} \beta_\varphi(\theta) d\lambda(\theta) \leqq \int_{\omega'} \beta_{\varphi'}(\theta)\, d\lambda(\theta)$$

also satisfies conditions (i) to (iv).

*Example* 4. Suppose that $\theta = (\theta_1, \cdots, \theta_r)$ and that $\omega$ consists of the single point $\theta^0 = (\theta_1^0, \cdots, \theta_r^0)$. We shall assume that the power function $\beta_\varphi(\theta)$ of any test $\varphi$ possesses continuous second derivatives $\partial^2\beta/\partial\theta_i\theta_j$ for all $i$ and $j$ at $\theta^0$. Let $\mathfrak{J}$ be the class of all level $\alpha$ tests that are strictly unbiased in the neighborhood of $\theta^0$ and let $\Delta(\varphi)$ denote the Gaussian curvature of the power surface at $\theta^0$, which is given by the determinant of the positive definite matrix $(\partial^2\beta/\partial\theta_i\partial\theta_j)\,|_{\theta^0}$. The relation $\varphi \lesssim \varphi'$ if $\Delta(\varphi) \leqq \Delta(\varphi')$ clearly satisfies (i) and (iii). It follows from a remark of Isaacson [1] that the relation is invariant provided the transformations $\bar{g}$ of the parameter space possess continuous second partial derivatives at $\theta^0$, which (under this restriction) verifies (iv). Condition (ii), finally, is easily verified. Optimum tests according to the present ordering correspond to the type D tests of Isaacson.

**2. Consequences of the Hunt-Stein theorem.** Under the assumptions of the preceding section we shall now show that if $G$ satisfies the conditions of the Hunt-Stein theorem (cf. [3], p. 336) and if there exists test $\psi_0$ which is optimum according to the ordering $\lesssim$, then there exists an almost invariant test which is optimum. Here we require of $\mathfrak{J}$ that it be closed under convex combinations and under weak limits.

The proof is completely analogous to and essentially follows from that of the Hunt-Stein theorem, and can be indicated very briefly. If $\nu_n$ is the sequence of almost invariant probability measures over $G$ postulated in the theorem, consider the sequence of tests

$$\psi_n = \int \psi_0 g\, d\nu_n(g)$$

Let $\psi$ be the weak limit of a subsequence $\psi_{n_i}$. Then it is shown in the proof of the Hunt-Stein theorem that $\psi$ is almost invariant, and it remains only to show that $\psi$ is optimum. By conditions (iv) and (ii) it follows for any $\varphi \varepsilon \mathfrak{J}$ that $\varphi \lesssim \psi_n$ for all $n$. Hence by condition (iii) also $\varphi \lesssim \psi$ for all $\varphi \varepsilon \mathfrak{J}$ as was to be proved.

Under the above assumptions, whenever there exists a UMP almost invariant test, this will be optimum with respect to any ordering $\lesssim$ satisfying conditions (i)–(iv). This explains the great variety of optimum properties possessed by certain tests and makes it unnecessary to prove each of them separately.

**3. Applications.** Consider a sequence of $n$ independent trials and let $X_i = 1$ or 0 as the $i$th trial is or is not successful. Let $P(X_i = 1) = p_i$ and consider the hypothesis $H: p_1 = \cdots = p_n = \frac{1}{2}$ against the alternatives

$$p_i > \tfrac{1}{2} \ (i = 1, \cdots, n).$$

The problem is invariant under any permutation of the variables and the sign test, which rejects when $\sum X_i > C$, is uniformly most powerful almost invariant (cf. [3], p. 219). This test therefore maximizes the minimum power over the alternatives $\omega'$: $\min p_i \geqq \frac{1}{2} + \Delta$ or $\omega$: $\max p_i \geqq \frac{1}{2} + \Delta$ for any $\Delta > 0$; it is most stringent and of type D.

As a second application, consider the general univariate linear hypothesis in the canonical form according to which the variables $X_1, \cdots, X_r$; $Y_1, \cdots, Y_s$; $Z_1, \cdots, Z_m$ are independently normally distributed with common variance $\sigma^2$ and means $E(X_i) = \xi_i$, $E(Y_j) = \eta_j$, $E(Z_k) = 0$. The hypothesis to be tested is $H: \xi_1 = \cdots = \xi_r = 0$. This problem remains invariant under the three groups

$G_1$: $Y_j' = Y_j + c_j(-\infty < c_j < \infty)$; $X_i' = X_i$; $Z_k' = Z_k$.

$G_2$: Orthogonal transformations of $X_1, \cdots, X_r$; $Y_j' = Y_j$; $Z_k' = Z_k$.

$G_3$: $X_i' = aX_i$; $Y_j' = aY_j$; $Z_k' = aZ_k$ $(a \neq 0)$.

The standard test has the following two basic optimum properties:

(a) It is uniformly most powerful among level $\alpha$ tests which are almost invariant with respect to $G_1, G_2, G_3$.

(b) It is uniformly most powerful among all unbiased (or similar) level $\alpha$ tests which are almost invariant with respect to $G_2$.

The first of these is well known; the second is easily shown by a standard argument.

Since the groups $G_1 - G_3$ satisfy the conditions of the Hunt-Stein theorem, it follows from (a), for example that the standard test is most stringent and that it maximizes the minimum power against the class of alternatives

$$\omega': \sum \xi_i^2/\sigma^2 \geqq \Delta.$$

To apply (b), consider fixed values of $\eta_1, \cdots, \eta_s$ and $\sigma$, so that the power becomes a function only of $\xi_1, \cdots, \xi_r$. It then follows that for any $\eta_1, \cdots, \eta_s$ and $\sigma$ the standard test maximizes (among all unbiased level $\alpha$ tests), for example

the minimum power over the sets $\omega'(\eta_1, \cdots, \eta_s, \sigma) : \sum \xi_i^2 \geq \Delta$ and the average power over the spheres $\sum \xi_i^2 = \Delta$. This was first proved by Wald [4]. It follows further that the test maximizes the Gaussian curvature of the power surface, considered for fixed $\eta_1, \cdots, \eta_s, \sigma$ as a function of the $\xi$'s, and hence is of Isaacson's type E. This has been shown previously by Kiefer [2], who deduced it as a consequence of the test maximizing the average power over the spheres $\sum \xi_i^2 = \Delta$.

## REFERENCES

[1] STANLEY L. ISAACSON, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 217–234.

[2] JACK KIEFER, "On the nonrandomized optimality and randomized nonoptimality of symmetrical designs," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 675–699.

[3] E. L. LEHMANN, *Testing Statistical Hypotheses*, New York, John Wiley and Sons, 1959.

[4] ABRAHAM WALD, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 434–439.

# THE WEIGHTED COMPOUNDING OF TWO INDEPENDENT SIGNIFICANCE TESTS

By M. Zelen and L. S. Joel

*National Bureau of Standards*

**1. Introduction and outline of the problem.** In a recent paper on the analysis of incomplete block designs [9], the situation arose where one had two statistically independent $F$ statistics for testing the same null hypothesis. A test was proposed for combining the two tests themselves into a single test which weighted one test relative to the other. It is the purpose of this paper to investigate numerically the power function of this proposed test as it will shed some light as to when an intra-block analysis is worthwhile.

Other situations where one has more than one independent test for testing the same null hypothesis are not uncommon. The tests may have arisen from several sets of independent data or from independent tests made on the same data. General discussions of combining independent tests can be found in Mosteller and Bush [4], Birnbaum [1], and E. S. Pearson [6]. For example a common situation in clinical experiments is that one desires to investigate the effects of two treatments (say) $t_1$ and $t_2$ on $2n + m$ people. It is known in advance that $m$ of these people will be available for receiving only one treatment. The experiment is run by assigning $t_1$ to $(m + n)$ subjects and $t_2$ to the remaining $n$ people. At a later time, $r$ new people are available who receive treatment $t_2$. Also of the $2n$ original remaining people, the $n$ people who first received $t_1$ receive $t_2$ and vice-versa. Thus the data consist of a cross-over design making use of $2n$ people, and also data where a person received only a single treatment. Thus it is possible to have two tests of the same null hypothesis that the treatments have no effect.[1]

The problem of combining information can be formulated as a problem in estimation. Generally for applications, this latter formulation is usually preferred as it will lead to confidence statements which are usually preferred to tests of a null hypothesis. However it seems interesting from a theoretical point of view to explore the consequences of combining the significance tests themselves.

Let there be two independent variance ratio statistics given by

$$F_j = s_{tj}^2/s_{ej}^2, \qquad j = 1, 2,$$

with degrees of freedom $\nu$ and $f_j$ $(j = 1, 2)$ respectively used to test the same null hypothesis. The numerator and denominator mean squares will be referred to as the "treatment" and "error" mean squares and are such that $f_j s_{ej}^2/\sigma_j^2$

Received February 20, 1958; revised January 8, 1959.

[1] We are indebted to Dr. S. Geisser, National Institute of Mental Health for pointing out this example.

follows a (central) chi-square distribution and $\nu s_{ij}^2/\sigma_j^2$ follows a non-central chi-square distribution with non-central parameter $\delta_j$, where $\delta_j$ is defined by

$$(1) \qquad \frac{E(s_{ij}^2)}{E(s_{ej}^2)} = 1 + \frac{\delta_j}{\nu}, \qquad j = 1, 2.$$

When $\delta_j = 0$, the null hypothesis is true and $\nu s_{ij}^2/\sigma_j^2$ will follow a (central) chi-square distribution.

For the purpose of combining the two tests, consider the integral transformation

$$(2) \qquad P_j = P\{F \geqq F_j \mid H_0\}, \qquad j = 1, 2,$$

which is the probability of the $F$-ratio exceeding the calculated $F_j$ if the null hypothesis is true. Our proposed method for combining the two tests is to use the critical region

$$(3) \qquad \omega:\{P_1 P_2^\theta \leqq C_\alpha\}$$

where $C_\alpha$ is a constant depending on an $\alpha$ level of significance and $\theta$ is a weighting factor $(0 \leqq \theta \leqq 1)$ which weights the second test relative to the first. (We will always assume that the first test has power which is equal to or greater than the second test.) This test is closely related to the procedure suggested by Good [3] for combining independent tests. Note that when $\theta = 0$ this corresponds to only using the first test; when $\theta = 1$, then both tests are given equal weight and the procedure is equivalent to the well-known method of Fisher [2] for combining independent tests of significance. The real problem here is to determine how to choose the weighting factor $\theta$. Our procedure for choosing $\theta$ is to let $\theta = \delta_2/\delta_1$, which in turn can be written as $\theta = (c_2/c_1)(\sigma_1^2/\sigma_2^2)$ where the $c_i$ are known constants. It is remarkable that this choice of a weighting factor results in minimum Type II error over a wide range of the other parameters involved.

**2. Distribution of the combined test.**

*Null distribution.* It is well known that when the null hypothesis is true, the distribution of $P_j$ will be that of a uniform random variable over the unit interval. Therefore the Type I error of the combined test is

$$(4) \qquad P(\omega \mid H_0) = P\{P_1 P_2^\theta \leqq C \mid H_0\} = \iint_\omega dP_1\, dP_2,$$

where $\omega$ denotes the critical region $\{P_1 P_2^\theta \leqq C\}$. Hence by an elementary integration we have

$$(5) \qquad P(\omega \mid H_0) = \begin{cases} C, & \text{for } \theta = 0, \\ \dfrac{C - \theta C^{1/\theta}}{1 - \theta} & \text{for } 0 < \theta < 1, \\ C(1 - \ln C), & \text{for } \theta = 1. \end{cases}$$

Therefore setting $P(\omega \mid H_0) = \alpha$ results in critical values of $C_\alpha$ for an $\alpha$ level of significance. Table I gives critical values of $C_\alpha$ for $\theta = 0.1(.1)1.0$ and $\alpha = .01, .05$.

TABLE I*

*Critical values of $C_\alpha$ for $\alpha = .05, .01$*

| $\theta$ | $\alpha = .05$ | $\alpha = .01$ |
|---|---|---|
| 0.0 | .050000 | .010000 |
| 0.1 | .045000 | .009000 |
| 0.2 | .040000 | .008000 |
| 0.3 | .035004 | .007000 |
| 0.4 | .030062 | .006001 |
| 0.5 | .025321 | .005013 |
| 0.6 | .020956 | .004062 |
| 0.7 | .017092 | .003190 |
| 0.8 | .013775 | .002432 |
| 0.9 | .010995 | .001805 |
| 1.0 | .008705 | .001309 |

$$*P\{P_1P_2^\theta \leq C_\alpha\} = \left\{\begin{matrix} C_\alpha \\ \dfrac{C_\alpha - \theta C_\alpha^{1/\theta}}{1-\theta} \\ C_\alpha(1 - \ln C_\alpha) \end{matrix}\right\} = \alpha, \quad \text{for} \quad \begin{matrix} \theta = 0 \\ 0 < \theta < 1 \\ \theta = 1 \end{matrix}$$

*Non-null distribution.* If

$$x_j = \frac{\nu F_j}{f_j + \nu F_j}, \tag{6}$$

then the non-null distribution of $x$ will have the p.d.f.

$$p(x_j \mid \delta_j) = e^{-\delta_j/2} \sum_{i=0}^{\infty} \frac{1}{B\left(\frac{\nu}{2} + i, \frac{f_j}{2}\right)} \frac{\delta_j^i}{2^i i!} x^{i+(\nu/2)-1} (1 - x_j)^{(f_j/2)-1} \quad (0 \leqq x_j \leqq 1) \tag{7}$$

and when $\delta_j = 0$, (7) reduces to the beta distribution,

$$p(x_j \mid 0) = \frac{1}{B\left(\frac{\nu}{2}, \frac{f_j}{2}\right)} x_j^{(\nu/2)-1}(1 - x_j)^{(f_j/2)-1}, \quad 0 \leqq x_j \leqq 1. \tag{8}$$

From the elementary properties of the probability integral transformation (cf. Pearson [6]) the p.d.f. of $P_j$ when the null hypothesis is not correct is given by

$$f(P_j) = \left. \frac{p(x_j \mid \delta_j)}{p(x_j \mid 0)} \right|_{x_j = g(P_j)} \quad (0 \leq P_j \leq 1), \tag{9}$$

where $x_j = g(P_j)$ means the solution of $x_j$ for a given value of $P_j$, where $x_j$ and $P_j$ are related by

$$P_j = \int_{x_j}^{1} p(x_j \mid 0)\, dx_j. \tag{10}$$

Hence substituting (7) and (8) in (9) results in the p.d.f. of the non-null distribution of $P_j$, i.e.,

$$f(P_j \mid \delta_j) = e^{-\delta_j/2} B\left(\frac{\nu}{2}, \frac{f_j}{2}\right) \sum_{i=0}^{\infty} \frac{1}{B\left(\frac{\nu}{2} + i, \frac{f_j}{2}\right)} \frac{\delta_j^i}{2^i i!} x_j^i, \tag{11}$$

where $x_j$ is related to $P_j$ by the incomplete beta function

$$P_j = I_{1-x_j}\left(\frac{f_j}{2}, \frac{\nu}{2}\right). \tag{12}$$

Therefore the power of the combined test for a given level of significance $\alpha$ is

$$P(\omega \mid H_1) = \iint_\omega f(P_1 \mid \delta_1) f(P_2 \mid \delta_2)\, dP_1\, dP_2, \tag{13}$$

where the region of integration is $\omega: \{P_1 P_2^\theta \leqq C_\alpha\}$.

The integral given in (13) is difficult to integrate as the p.d.f. $f(P_j \mid \delta_j)$ is not an explicit function of $P_j$. In order to evaluate (13) numerically it is convenient to consider the integral transformation

$$\pi_j = \int_{P_j}^1 f(P \mid \delta_j)\, dP, \qquad j = 1, 2. \tag{14}$$

Then (13) can be written

$$P(\omega \mid H_1) = \iint_{\omega^*} d\pi_1\, d\pi_2 \tag{15}$$

where $\omega^*$ denotes the region $\omega$ in terms of $\pi_1$ and $\pi_2$. Thus to every point on the boundary $P_1 P_2^\theta = C_\alpha$ in the $(P_1, P_2)$ space there will correspond a point in the $(\pi_1, \pi_2)$ space and it will be possible to map the region $\omega^*$ entirely, even though we do not have an explicit expression in the $\pi_1$, $\pi_2$ variables for the boundary.

For this purpose it is convenient to find $\pi_j$ from the non-central distribution of $x_j$, i.e.,

$$\pi_j = \int_{P_j}^1 f(P \mid \delta_j)\, dP = \int_{x_j}^1 p(x \mid \delta_j)\, dx, \qquad j = 1, 2. \tag{16}$$

Unfortunately the non-central distribution given by (16) is only tabulated for values of $x_j$ corresponding to $P_j = .01$ and .05. However it is possible to use the Patnaik approximation to the non-central $F$ (or equivalent beta) distribution [5] and find approximate values for $\pi_j$. (This approximation appears to have a maximum error of one unit in the second decimal.) For purposes of tabulation, it is more convenient to use the non-central variable $\Delta_j = \delta_j/\nu$ which is related to Tang's non-central parameter $\Phi$, [7], by $\Phi = [\nu\Delta/(\nu + 1)]^{\frac{1}{2}}$. Then in terms of $\Delta_j$, the Patnaik approximation can be written

$$\int_0^{x_j} p(x \mid \delta_j)\, dx \approx I_{x_j'}\left(\frac{n}{2}, \frac{f_j}{2}\right) \tag{17}$$

where

$$(18)\qquad \begin{cases} n = \dfrac{\nu(1 + \Delta_j)^2}{(1 + 2\Delta_j)} \\ x_j' = \dfrac{(1 + \Delta_j)x_j}{1 + 2\Delta_j - \Delta_j x_j}. \end{cases}$$

Table II summarizes calculations for the Type II error ($P_{II}$) for the parameters

$$\alpha = .05, \theta = 0(.2)1.0, \Delta_1 = 1(2)5, \Delta_2 = 0(1)4, \Delta_2 < \Delta_1,$$

$$(\nu, f_1, f_2) = (5, 10, 5), (5, 15, 5), (5, 15, 10), (10, 10, 5), (10, 15, 5), (10, 15, 10), (5, 30, 10), (5, 30, 15), (5, 30, 20), (5, 30, 25).$$

$$\alpha = .05, \theta = 0(.2)1.0, \Delta_1 = 1(2)3, \Delta_2 = 0(1)2, \Delta_2 < \Delta_1,$$

$$\nu = 10, 15, f_1 = 30, f_2 = 10(5)25.$$

$$\alpha = .01, \theta = 0(.2)1.0, \Delta_1 = 1(2)5, \Delta_2 = 0(1)4, \Delta_2 < \Delta_1, \nu = 5, 10$$

$$(f_1, f_2) = (10, 5), (15, 5), (15, 10), f_1 = 30, f_2 = 10(5)25.$$

$$\alpha = .01, \theta = 0(.2)1.0, \Delta_1 = 1(2)5, \Delta_2 = 0(1)4, \Delta_2 < \Delta_1, \nu = 15$$

$$(f_1, f_2) = (15, 5), (15, 10), f_1 = 30, f_2 = 10(5)25.$$

$$\alpha = .01, \theta = 0(.2)1.0, \Delta_1 = 7, \Delta_2 = 0(2)6, \nu = 5, 10$$

$$(f_1, f_2) = (10, 5), (15, 5), (15, 10).$$

Since Patnaik's approximation may be in error by one unit in the second decimal place, the accuracy of Table II is limited to at best an error of the same magnitude. Interpolation in the table on any of the degrees of freedom parameters should be made using the reciprocals, i.e., $\nu^{-1}, f_j^{-1}$.

**3. The effect of the weight factor on the Type II error.** A typical Type II error curve is graphed in Fig. 1 for the parameters $\alpha = .05$, $\nu = 5$, $f_1 = 10$, $f_2 = 5$. Note that it is possible for the Type II error of the combined test to be larger than if a single test had been used alone. This corresponds to the case when the second test is given too much weight.

Note also that the minimum $P_{II}$ is rather flat. For example for $\Delta_1 = 5$, $\Delta_2 = 1$ the minimum is between $\theta = .2$ and $\theta = .4$. This is typical of the behavior of $P_{II}$. Table III shows the range of $\theta$ for which the minimum $P_{II}$ (to two decimal places) was attained. Also given in this table is the ratio $\Delta_2/\Delta_1 = \delta_2/\delta_1$ which we put forward as the weighting factor. In the entire table of $P_{II}$ this choice of $\theta$ will result in being off by at most one unit in the second decimal from the minimum $P_{II}$.

TABLE II: *Values of* $P_{II}$ *(Type* II *error) for* $\alpha$ = *.05*

| $\nu$ | $\Delta_1$ | $\Delta_2$ | $f_1 =$ | 10 | | | | | | 15 | | | | | | 15 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $f_2 =$ | 5 | | | | | | 5 | | | | | | 10 | | | | | |
| | | | $\theta =$ | 0.0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 |
| 5 | 1 | 0 | | .793 | .807 | .817 | .828 | .842 | .856 | .779 | .783 | .793 | .807 | .823 | .839 | | | | | | |
| | 3 | 0 | | .405 | .410 | .429 | .465 | .512 | .562 | .322 | .327 | .346 | .380 | .424 | .474 | | | | | | |
| | | 1 | | | .381 | .373 | .384 | .410 | .442 | | .302 | .297 | .309 | .334 | .366 | .322 | .294 | .283 | .290 | .310 | .338 |
| | | 2 | | | .345 | .306 | .292 | .294 | .307 | | .272 | .241 | .231 | .235 | .248 | | .251 | .206 | .187 | .185 | .191 |
| | 5 | 0 | | .150 | .155 | .173 | .206 | .252 | .306 | .087 | .090 | .103 | .125 | .158 | .200 | | | | | | |
| | | 1 | | | .137 | .138 | .154 | .181 | .215 | | .079 | .080 | .092 | .111 | .137 | .087 | .076 | .075 | .084 | .101 | .124 |
| | | 2 | | | .116 | .101 | .102 | .112 | .129 | | .066 | .058 | .059 | .067 | .079 | | .058 | .047 | .046 | .050 | .058 |
| | | 3 | | | .105 | .083 | .078 | .082 | .092 | | .059 | .047 | .045 | .048 | .055 | | .049 | .034 | .030 | .031 | .035 |
| | | 4 | | | .096 | .070 | .061 | .062 | .067 | | .054 | .039 | .035 | .036 | .040 | | .042 | .025 | .020 | .020 | .021 |
| 10 | 1 | 0 | | .741 | .744 | .755 | .772 | .793 | .814 | .690 | .692 | .704 | .724 | .750 | .776 | | | | | | |
| | 3 | 0 | | .294 | .299 | .321 | .360 | .411 | .467 | .184 | .188 | .206 | .238 | .282 | .333 | | | | | | |
| | | 1 | | | .264 | .253 | .263 | .286 | .317 | | .163 | .158 | .167 | .187 | .215 | .184 | .153 | .140 | .144 | .158 | .179 |
| | | 2 | | | .235 | .200 | .190 | .195 | .208 | | .143 | .121 | .117 | .123 | .135 | | .122 | .091 | .080 | .080 | .085 |
| | 5 | 0 | | .071 | .075 | .088 | .114 | .151 | .199 | .023 | .025 | .030 | .042 | .060 | .085 | | | | | | |
| | | 1 | | | .061 | .061 | .070 | .088 | .112 | | .019 | .020 | .024 | .032 | .044 | .023 | .018 | .017 | .020 | .026 | .036 |
| | | 2 | | | .050 | .042 | .043 | .051 | .062 | | .016 | .013 | .014 | .018 | .023 | | .012 | .009 | .009 | .011 | .014 |
| | | 3 | | | .043 | .032 | .031 | .034 | .040 | | .013 | .010 | .010 | .012 | .015 | | .009 | .006 | .005 | .006 | .007 |
| | | 4 | | | .037 | .024 | .021 | .022 | .026 | | .011 | .007 | .007 | .007 | .009 | | .007 | .003 | .003 | .003 | .004 |

| $\nu$ | $\Delta_1$ | $\Delta_2$ | $f_1 =$ | 30 | | | | | | 30 | | | | | | 30 | | | | | | 30 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $f_2 =$ | 10 | | | | | | 15 | | | | | | 20 | | | | | | 25 | | | | |
| | | | $\theta =$ | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 |
| 5 | 1 | 0 | .746 | .752 | .762 | .777 | .796 | .814 | | | | | | | | | | | | | | | | | | |
| | 3 | 0 | .235 | .239 | .255 | .284 | .322 | .367 | | | | | | | | | | | | | | | | | | |
| | | 1 | | .220 | .217 | .228 | .250 | .278 | .235 | .213 | .206 | .213 | .231 | .256 | .235 | .210 | .202 | .206 | .223 | .246 | .235 | .208 | .197 | .202 | .218 | .240 |
| | | 2 | | .196 | .174 | .167 | .172 | .184 | | .180 | .148 | .135 | .135 | .141 | | .172 | .135 | .120 | .118 | .123 | | .167 | .127 | .112 | .109 | .113 |
| | 5 | 0 | .040 | .042 | .048 | .060 | .079 | .103 | | | | | | | | | | | | | | | | | | |
| | | 1 | | .036 | .037 | .043 | .054 | .068 | .040 | .034 | .034 | .040 | .049 | .062 | .040 | .034 | .033 | .038 | .047 | .059 | .040 | .033 | .033 | .037 | .046 | .058 |
| | | 2 | | .030 | .026 | .027 | .031 | .038 | | .026 | .021 | .021 | .024 | .028 | | .024 | .019 | .018 | .020 | .024 | | .023 | .017 | .017 | .019 | .022 |
| | | 3 | | .027 | .021 | .020 | .022 | .026 | | .022 | .015 | .014 | .014 | .016 | | .019 | .012 | .011 | .011 | .013 | | .018 | .011 | .010 | .010 | .011 |
| | | 4 | | .024 | .017 | .016 | .016 | .019 | | .018 | .011 | .009 | .009 | .010 | | .016 | .008 | .007 | .007 | .007 | | .014 | .007 | .006 | .005 | .006 |
| 10 | 1 | 0 | .608 | .613 | .626 | .650 | .680 | .711 | | | | | | | | | | | | | | | | | | |
| | 3 | 0 | .082 | .086 | .096 | .115 | .142 | .177 | | | | | | | | | | | | | | | | | | |
| | | 1 | | .073 | .071 | .077 | .090 | .108 | .082 | .068 | .062 | .066 | .075 | .088 | .082 | .065 | .058 | .060 | .068 | .079 | .082 | .063 | .055 | .056 | .063 | .074 |
| | | 2 | | .062 | .053 | .052 | .056 | .064 | | .053 | .039 | .035 | .036 | .040 | | .047 | .032 | .027 | .027 | .029 | | .043 | .027 | .023 | .022 | .024 |
| 15 | 1 | 0 | .518 | .521 | .537 | .565 | .601 | .639 | | | | | | | | | | | | | | | | | | |
| | 3 | 0 | .029 | .030 | .036 | .046 | .062 | .084 | | | | | | | | | | | | | | | | | | |
| | | 1 | | .024 | .024 | .027 | .034 | .043 | .029 | .022 | .019 | .021 | .026 | .032 | .029 | .020 | .017 | .018 | .022 | .027 | .029 | .019 | .016 | .016 | .019 | .024 |
| | | 2 | | .020 | .017 | .017 | .019 | .024 | | .016 | .011 | .010 | .011 | .013 | | .013 | .008 | .007 | .007 | .008 | | .011 | .006 | .005 | .006 | .006 |

## TABLE II (*Continued*): *Values of* $P_{II}$ (*Type* II *error*) *for* $\alpha = .01$

| $\nu$ | $\Delta_1$ | $\Delta_2$ | $f_1 =$ 10 | | | | | | 15 | | | | | | 15 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $f_2 =$ 5 | | | | | | 5 | | | | | | 10 | | | | | |
| | | | $\theta =$ 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 |
| 5 | 1 | 0 | .921 | .923 | .930 | .935 | .940 | .944 | .904 | .911 | .919 | .925 | .932 | .938 | | | | | | |
| | 3 | 0 | .714 | .720 | .732 | .755 | .788 | .824 | .622 | .628 | .641 | .668 | .708 | .754 | | | | | | |
| | | 1 | | .705 | .692 | .694 | .713 | .742 | | .610 | .598 | .604 | .628 | .664 | .622 | .602 | .583 | .582 | .600 | .632 |
| | | 2 | | .680 | .636 | .612 | .611 | .626 | | .582 | .542 | .523 | .526 | .546 | | .560 | .500 | .462 | .452 | .461 |
| | 5 | 0 | .455 | .459 | .477 | .514 | .570 | .638 | .310 | .314 | .331 | .366 | .421 | .491 | | | | | | |
| | | 1 | | .434 | .426 | .440 | .477 | .529 | | .293 | .289 | .305 | .340 | .393 | .310 | .287 | .277 | .288 | .319 | .366 |
| | | 2 | | .402 | .364 | .353 | .367 | .400 | | .268 | .242 | .236 | .252 | .283 | | .250 | .211 | .197 | .204 | .226 |
| | | 3 | | .383 | .331 | .306 | .309 | .330 | | .254 | .216 | .201 | .207 | .228 | | .229 | .176 | .151 | .148 | .159 |
| | | 4 | | .368 | .303 | .268 | .262 | .276 | | .242 | .195 | .173 | .172 | .186 | | .211 | .147 | .117 | .109 | .113 |
| | 7 | 0 | .248 | .253 | .271 | .310 | .372 | .452 | .122 | .125 | .138 | .164 | .209 | .272 | | | | | | |
| | | 2 | | .206 | .183 | .182 | .201 | .237 | | .098 | .087 | .088 | .102 | .128 | .122 | .089 | .073 | .071 | .077 | .098 |
| | | 4 | | .181 | .140 | .124 | .128 | .146 | | .084 | .064 | .058 | .062 | .074 | | .069 | .044 | .036 | .037 | .043 |
| | | 6 | | .160 | .107 | .084 | .080 | .086 | | .073 | .047 | .037 | .037 | .042 | | .053 | .026 | .018 | .017 | .018 |
| 10 | 1 | 0 | .896 | .903 | .912 | .919 | .926 | .934 | .868 | .877 | .886 | .895 | .907 | .919 | | | | | | |
| | 3 | 0 | .629 | .637 | .651 | .680 | .722 | .771 | .474 | .482 | .498 | .531 | .583 | .645 | | | | | | |
| | | 1 | | .608 | .588 | .587 | .609 | .645 | | .451 | .436 | .441 | .468 | .512 | .474 | .437 | .408 | .402 | .420 | .455 |
| | | 2 | | .579 | .529 | .503 | .504 | .524 | | .423 | .382 | .365 | .372 | .398 | | .392 | .325 | .288 | .280 | .291 |
| | 5 | 0 | .322 | .327 | .346 | .387 | .450 | .529 | .148 | .152 | .166 | .196 | .245 | .314 | | | | | | |
| | | 1 | | .295 | .284 | .296 | .332 | .385 | | .133 | .129 | .140 | .167 | .209 | .148 | .125 | .116 | .122 | .142 | .177 |
| | | 2 | | .267 | .234 | .226 | .241 | .274 | | .118 | .102 | .100 | .113 | .138 | | .102 | .078 | .071 | .076 | .091 |
| | | 3 | | .250 | .204 | .185 | .190 | .212 | | .108 | .086 | .079 | .085 | .101 | | .088 | .057 | .047 | .047 | .054 |
| | | 4 | | .234 | .176 | .148 | .146 | .158 | | .099 | .072 | .061 | .062 | .072 | | .075 | .041 | .029 | .028 | .030 |
| | 7 | 0 | .133 | .137 | .153 | .187 | .243 | .321 | .032 | .034 | .040 | .053 | .078 | .119 | | | | | | |
| | | 2 | | .101 | .087 | .087 | .102 | .130 | | .023 | .020 | .021 | .027 | .039 | | | | | | |
| | | 4 | | .083 | .057 | .049 | .053 | .064 | | .018 | .012 | .011 | .013 | .018 | | | | | | |
| | | 6 | | .070 | .040 | .030 | .030 | .035 | | .014 | .008 | .006 | .007 | .010 | | | | | | |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 1 | 0 | | | | | | | .842 | .854 | .864 | .875 | .890 | .906 | | | | | | |
| | 3 | 0 | | | | | | | .372 | .376 | .393 | .430 | .487 | .557 | | | | | | |
| | | 1 | | | | | | | | .343 | .328 | .334 | .363 | .409 | .372 | .326 | .297 | .291 | .309 | .344 |
| | | 2 | | | | | | | | .317 | .280 | .266 | .277 | .304 | | .284 | .222 | .192 | .188 | .199 |
| | 5 | 0 | | | | | | | .082 | .085 | .096 | .119 | .159 | .219 | | | | | | |
| | | 1 | | | | | | | | .071 | .069 | .077 | .097 | .129 | .082 | .065 | .059 | .062 | .076 | .101 |
| | | 2 | | | | | | | | .061 | .052 | .052 | .062 | .080 | | .050 | .036 | .033 | .037 | .046 |
| | | 3 | | | | | | | | .054 | .041 | .038 | .042 | .053 | | .040 | .023 | .019 | .020 | .024 |
| | | 4 | | | | | | | | .049 | .033 | .028 | .030 | .037 | | .033 | .016 | .012 | .012 | .014 |

| | | | $f_1 =$ | 30 | | | | | | 30 | | | | | | 30 | | | | | | 30 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r$ | $\Delta_1$ | $\Delta_2$ | $f_2 =$ | 10 | | | | | | 15 | | | | | | 20 | | | | | | 25 | | | | | |
| | | | $\theta =$ | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 | 0 | .2 | .4 | .6 | .8 | 1.0 |
| 5 | 1 | 0 | | .883 | .892 | .901 | .909 | .918 | .927 | | | | | | | | | | | | | | | | | | |
| | 3 | 0 | | .500 | .502 | .515 | .543 | .587 | .640 | | | | | | | | | | | | | | | | | | |
| | | 1 | | | .483 | .474 | .483 | .510 | .550 | .500 | .476 | .461 | .463 | .485 | .521 | .500 | .472 | .453 | .453 | .473 | .507 | .500 | .469 | .449 | .447 | .465 | .498 |
| | | 2 | | | .457 | .424 | .409 | .416 | .440 | | .438 | .387 | .358 | .354 | .367 | | .427 | .367 | .332 | .322 | .331 | | .420 | .353 | .315 | .303 | .310 |
| | 5 | 0 | | .166 | .169 | .181 | .206 | .246 | .303 | | | | | | | | | | | | | | | | | | |
| | | 1 | | | .156 | .155 | .166 | .193 | .233 | .166 | .152 | .148 | .156 | .180 | .216 | .166 | .150 | .144 | .152 | .173 | .208 | .166 | .149 | .142 | .149 | .170 | .203 |
| | | 2 | | | .140 | .126 | .125 | .137 | .160 | | .130 | .109 | .103 | .109 | .126 | | .124 | .100 | .092 | .097 | .111 | | .121 | .095 | .086 | .090 | .102 |
| | | 3 | | | .132 | .111 | .104 | .110 | .125 | | .118 | .089 | .077 | .077 | .086 | | .110 | .077 | .065 | .064 | .070 | | .104 | .071 | .058 | .056 | .061 |
| | | 4 | | | .125 | .099 | .088 | .090 | .100 | | .107 | .073 | .058 | .056 | .060 | | .097 | .060 | .046 | .042 | .045 | | .091 | .053 | .039 | .035 | .037 |
| 10 | 1 | 0 | | .819 | .825 | .836 | .849 | .867 | .886 | | | | | | | | | | | | | | | | | | |
| | 3 | 0 | | .270 | .273 | .287 | .316 | .364 | .426 | | | | | | | | | | | | | | | | | | |
| | | 1 | | | .250 | .242 | .250 | .276 | .316 | .270 | .240 | .224 | .224 | .243 | .275 | .270 | .234 | .213 | .210 | .225 | .253 | .270 | .230 | .206 | .200 | .213 | .240 |
| | | 2 | | | .230 | .205 | .198 | .208 | .231 | | .210 | .170 | .151 | .150 | .163 | | .196 | .148 | .126 | .121 | .129 | | .187 | .134 | .110 | .104 | .109 |
| | 5 | 0 | | .031 | .032 | .037 | .047 | .064 | .093 | | | | | | | | | | | | | | | | | | |
| | | 2 | | | .024 | .020 | .021 | .025 | .034 | | | | | | | | | | | | | | | | | | |
| | | 4 | | | .019 | .013 | .012 | .013 | .016 | | | | | | | | | | | | | | | | | | |

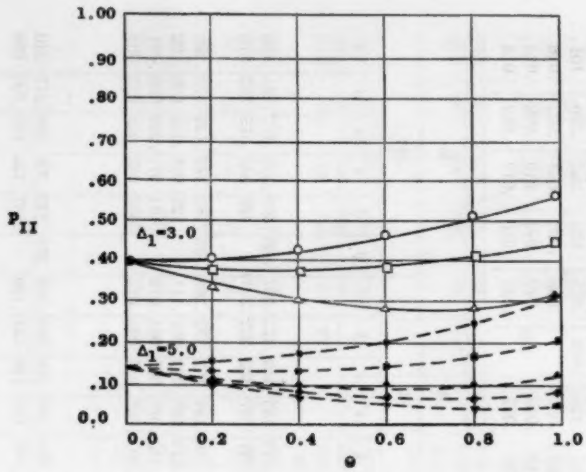


FIG. 1. Type II error ($P_{II}$) for $\nu = 5, f_1 = 10, f_2 = 5, \alpha = .05$.

**TABLE III**: *Range of $\theta$ for which minimum $P_{II}$ is attained for $\nu = 5, 10$ and $\alpha = .01, .05$*

| $\alpha$ | $\nu - f_1 - f_2$ | $\Delta_1 =$ | 3 | | 5 | | | | 7 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\Delta_2 =$ | 1 | 2 | 1 | 2 | 3 | 4 | 2 | 4 | 6 |
| | | $\frac{\Delta_2}{\Delta_1} =$ | .33 | .67 | .20 | .40 | .60 | .80 | .29 | .57 | .86 |
| .01 | 5-10- 5 | | .4-.6 | .6-.8 | .2-.4 | .6 | .6-.8 | .8 | .4-.6 | .6 | .6-.8 |
| | 15- 5 | | .4-.6 | .6 | .2-.4 | .4-.6 | .6 | .6-.8 | .4-.6 | .6-.8 | .6-.8 |
| | 10 | | .4-.6 | .8 | .4 | .6-.8 | .6-.8 | .8-1.0 | .4-.6 | .6-.8 | .6-.8 |
| | 10-10- 5 | | .4-.6 | .6-.8 | .4 | .4-.6 | .6 | .6-.8 | .4-.6 | .6-.8 | .6-.8 |
| | 15- 5 | | .4-.6 | .6 | .2-.4 | .4-.6 | .6 | .6-.8 | .4-.6 | .4-.6 | .6-.8 |
| | 10 | | .6 | .8 | .4-.6 | .6 | .6-.8 | .6-.8 | .4-.6 | .4-.6 | .4-.6 |
| | 5-30-10 | | .4 | .6-.8 | .2-.4 | .4 | .6 | .4-.8 | | | |
| | 15 | | .4-.6 | .8 | .2-.4 | .6 | .6-.8 | .6-1.0 | | | |
| | 20 | | .4-.6 | .8 | .4 | .6 | .6-.8 | .8-1.0 | | | |
| | 25 | | .4-.6 | .8 | .4 | .6-.8 | .6-1.0 | .6-1.0 | | | |
| | 10-30-10 | | .4 | .4-.6 | | .2-.8 | | .4-.8 | | | |
| | 15 | | .4-.6 | .6-.8 | | | | | | | |
| | 20 | | .4-.6 | .8 | | | | | | | |
| | 25 | | .6 | .8 | | | | | | | |
| .05 | 5-10- 5 | | .4 | .6-.8 | .2-.4 | .4-.6 | .4-.6 | .6-.8 | | | |
| | 15- 5 | | .2-.4 | .6 | .2-.4 | .4-.6 | .6 | .4-1.0 | | | |
| | 10 | | .4 | .6-1.0 | .2-.6 | .4-.8 | .4-.8 | .6-1.0 | | | |
| | 10-10- 5 | | .4 | .6-.8 | .2-.4 | .4-.6 | .4-.8 | .4-1.0 | | | |
| | 15- 5 | | .2-.4 | .4-.8 | .2-.4 | .4-.6 | .2-.8 | .2-1.0 | | | |
| | 10 | | .4-.6 | .6-.8 | .2-.6 | .2-1.0 | .6 | .4-1.0 | | | |
| | 5-30-10 | | .2-.4 | .4-.8 | .2-.4 | .2-.8 | .4-.8 | .2-1.0 | | | |
| | 15 | | .2-.6 | .6-1.0 | .2-.4 | .4-.8 | .6-.8 | .4-1.0 | | | |
| | 20 | | .2-.6 | .6-1.0 | .2-.4 | .2-1.0 | .4-1.0 | .4-1.0 | | | |
| | 25 | | .4-.6 | .6-1.0 | .2-.4 | .2-1.0 | .4-1.0 | .2-1.0 | | | |
| | 10-30-10 | | .2-.4 | .4-.6 | | | | | | | |
| | 15 | | .4 | .4-1.0 | | | | | | | |
| | 20 | | .4-.6 | .4-1.0 | | | | | | | |
| | 25 | | .2-.8 | .6-1.0 | | | | | | | |

In general the non-centrality parameter can be written as

$$\delta_j = \frac{c_j}{\sigma_j^2}\mu^2, \qquad j = 1, 2, \tag{19}$$

where $c_j$ is a known constant which depends on how the observations were taken, $\sigma_j^2$ is an underlying population variance, and $\mu^2$ is a non-negative constant which depends on the particular hypothesis involved and is only equal to zero if the null hypothesis is true. Hence,

$$\theta = \frac{\delta_2}{\delta_1} = \frac{c_2}{c_1}\frac{\sigma_1^2}{\sigma_2^2}, \tag{20}$$

which is a function only of the known constants $c_j$ and the ratio of the population variances. Of course in many practical situations the ratio of the variances $\sigma_1^2/\sigma_2^2$ may not be known. In this case we believe that the estimate for $\sigma_1^2/\sigma_2^2$ can be used in the weighting factor. This recommendation is based on the fact that the weighting factor need not be known accurately in order to achieve a minimum $P_{II}$. However it should be pointed out that this latter procedure will result in a change in the significance level and power of the test.

## REFERENCES


[1] A. Birnbaum, "Combining independent tests of significance," *J. Am. Stat. Assoc.*, Vol. 49 (1954), pp. 559–574.

[2] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, Ltd., Edinburgh, 1938.

[3] I. J. Good, "On the weighted combination of significance tests," *J.R.S.S.*, Ser. B, Vol. 17 (1955), pp. 264–265.

[4] F. Mosteller and R. R. Bush, "Selected quantitative techniques," *Handbook of Social Psychology*, Chapter 8, Addison-Wesley, Cambridge, Mass., 1954.

[5] P. B. Patnaik, "The non-central $\chi^2$ and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[6] E. S. Pearson, "The probability integral transformation for testing goodness of fit and combining independent tests of significance," *Biometrika*, Vol. 30 (1938), pp. 134–148.

[7] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 126–149.

[8] W. A. Wallis, "Compounding probabilities from independent significance tests," *Econometrica*, Vol. 10 (1942), pp. 229–248.

[9] M. Zelen, "The analysis of incomplete block designs," *J. Am. Stat. Assoc.*, Vol. 52 (1957), pp. 204–217.

# BAYES ACCEPTANCE SAMPLING PROCEDURES FOR LARGE LOTS[1]

By D. Guthrie, Jr. and M. V. Johns, Jr.

*Stanford University*

**1. Introduction and statement of the main results.** A lot consisting of $N$ items may be characterized by $N$ non-negative random variables $X_i$, $i = 1, 2, \cdots, N$, where the value of $X_i$ indicates the quality of the $i$th item. In a typical case $X_i$ might take on the values zero and one according to whether the $i$th item is non-defective or defective. Alternatively, $X_i$ might be defined to be the number of defects in the $i$th item so that the possible values of $X_i$ would be $0, 1, 2, \cdots$. In still another formulation $X_i$ might be a continuous random variable related to the deviation from standard of some characteristic of the item. We shall assume that the random variables $X_i$, $i = 1, 2, \cdots, N$, are independent and identically distributed with common distribution function $F(x \mid \lambda)$ depending on a single parameter $\lambda$.

The fixed size sampling inspection scheme to be considered consists of the random selection of $n$ items from the lot and the observation of the values of the corresponding $X_i$'s. Thus, the sample may be described by the random variables $X_1, X_2, \cdots, X_n$. The two possible actions to be taken on the basis of the sample are acceptance or rejection of the uninspected remainder of the lot. The consequences of these alternative actions are appraised by the following cost model where we let $S_k = \sum_{i=1}^{k} X_i$ for any $k = 1, 2, \cdots, N$:

| | Action | Cost |
|---|---|---|
| (1.1) | Acceptance | $a_1(S_N - S_n) + a_2(N - n) + s_1S_n + s_2n$ |
| | Rejection | $r_1(S_N - S_n) + r_2(N - n) + s_1S_n + s_2n.$ |

Thus, for $i = n + 1, n + 2, \cdots, N$, the contributions to the total cost due to the acceptance or rejection of the $i$th item without inspection are given by $a_1X_i + a_2$ and $r_1X_i + r_2$ respectively. For $i = 1, 2, \cdots, n$ the cost of inspection (and possibly replacement) of the $i$th item is given by $s_1X_i + s_2$. If, for example, an item is classified as defective or non-defective by $X_i$, then $S_N$ and $S_n$ are the number of defective items in the lot and in the sample respectively. Suppose that the cost of accepting an item is $a_1$ if the item is defective and zero if the item is non-defective, and that the cost of rejecting the uninspected remainder of the lot is proportional to the number of items remaining in the lot. Then $a_2 = r_1 = 0$. If, in addition, all items found to be defective in the sample are replaced with good items, each at a cost of $s_1$ units, and $s_2$ represents the cost of the time and labor required to inspect each item in the sample, then (1.1)

Received June 16, 1958.

[1] Prepared for Army, Navy, and Air Force under Contract N6onr25126 with the Office of Naval Research. Reproduction in whole or part permitted for any U. S. Government purpose.

becomes

| | Action | Cost |
|---|---|---|
| (1.1a) | Acceptance | $a_1(S_N - S_n) + s_1S_n + s_2n$ |
| | Rejection | $r_2(N - n) + s_1S_n + s_2n.$ |

The cost model (1.1) includes a wide variety of sampling inspection and acceptance sampling problems corresponding to various choices of the cost parameters and the family of distribution functions $F(x \mid \lambda)$. Similar formulations of this problem have been given by several authors, notably [1], [2], and [3].

The authors of [1] and [2] have attempted to characterize optimal sample sizes in terms of a minimax criterion. When the lot size is large this approach seems to lead to sample sizes which are appropriate when the true state of nature (value of $\lambda$) has a high *a priori* probability of being very close to the "indifference state" where either acceptance or rejection leads to the same expected cost. Such an *a priori* assumption about the true state of nature will not generally be reasonable, which suggests that the minimax criterion is not suitable for this problem.

The purpose of this paper is to find explicit asymptotic characterizations for large $N$ of the decision procedures and sample sizes which are optimal in the Bayes sense for various classes of *a priori* probability distributions defined over the values of the parameter $\lambda$. This problem is considered for certain families of distribution functions $F(x \mid \lambda)$ of the exponential type having the property that $E(X \mid \lambda) = \lambda$. This parametrization of distribution functions of the exponential type is convenient because (1) for this case the range of possible values of $X$ coincides with the range of $\lambda$, and (2) any available *a priori* information will usually be most easily expressed in terms of the expected quality of an item, i.e., the value of the parameter $\lambda$.

The two principal reasons for investigating the Bayes solutions to this problem are as follows: (1) In most practical situations the statistician will possess some subjective *a priori* information concerning the probable values of the parameter $\lambda$ and such information may often be reasonably summarized and made objective by the choice of a suitable *a priori* distribution; (2) for statistical decision problems of the type under consideration, the class of Bayes decision procedures coincides with the admissible class so that all procedures discussed will have the optimal property of admissibility (see, e.g., [4]).

Each family of distribution functions $F(x \mid \lambda)$ to be considered will be defined in terms of a given measure $\mu$ on the Borel sets of the positive real half-line as follows: Let

$$(1.2)\qquad \begin{aligned} a &= \inf\left\{x: \int_x^{x+\epsilon} d\mu > 0, \quad \text{all} \quad \epsilon > 0\right\}, \\ b &= \sup\left\{x: \int_{x-\epsilon}^{x} d\mu > 0, \quad \text{all} \quad \epsilon > 0\right\}. \end{aligned}$$

Let $I_\mu$ be the interval $[a, b]$ if $b < \infty$ and $[a, \infty)$ if $b = \infty$. We assume that $\mu$ satisfies the conditions

(1.3)
A) $I_\mu$ is non-empty,
B) There exists a function $\omega(\lambda)$ such that for each $\lambda \varepsilon I_\mu$

$$\frac{\int_{[a,\infty)} xe^{\omega(\lambda)x}\, d\mu(x)}{\int_{[a,\infty)} e^{\omega(\lambda)x}\, d\mu(x)} = \lambda.$$

We now define $F(x \mid \lambda)$ by

$$F(x \mid \lambda) = \begin{cases} 0, & x \leqq a, \\ \dfrac{\int_{[a,x)} e^{\omega(\lambda)t}\, d\mu(t)}{\int_{[a,\infty)} e^{\omega(\lambda)t}\, d\mu(t)}, & a < x \leqq b, \\ 1, & x > b. \end{cases} \tag{1.4}$$

The following theorems concerning such families are proved in Section 2.

THEOREM 2.1: *The function $\omega(\lambda)$ given by (1.3B) is unique and $d\omega(\lambda)/d\lambda$ exists and is positive for $\lambda \varepsilon I_\mu$.*

THEOREM 2.2: *If $F(x \mid \lambda)$ is defined by (1.4), then all moments of $F(x \mid \lambda)$ exist and all derivatives of $\omega(\lambda)$ exist and are finite for $\lambda \varepsilon I_\mu$.*

THEOREM 2.3: *The distribution function $F(x \mid \lambda)$ defined by (1.4) may be represented for $a < x \leqq b$ and for $\lambda \varepsilon I_\mu$ by*

$$F(x \mid \lambda) = K(\gamma) \int_{[a,x)} \exp\left\{\omega(\lambda)t - \int_\gamma^\lambda u\omega'(u)\, du\right\} d\mu(t), \tag{1.5}$$

*if and only if assumption (1.3B) is satisfied, where $\gamma \varepsilon I_\mu$ and $K(\gamma)$ is a normalizing factor depending on the choice of $\gamma$ and determined so that $F(b+ \mid \lambda) = 1$.*

We may unambiguously define $F(x \mid a)$ and, if $b$ is finite, $F(x \mid b+)$ by

$$F(x \mid a) = \lim_{\lambda \searrow a} F(x \mid \lambda), \tag{1.6}$$

and

$$F(x \mid b+) = \lim_{\lambda \nearrow b} F(x \mid \lambda). \tag{1.7}$$

It is easily verified that the $n$-fold convolution of $F(x \mid \lambda)$ may be written

$$F^{(n)}(x \mid \lambda) = (K(\gamma))^n \int_{[na,x)} \exp\left\{\omega(\lambda)t - n\int_\gamma^\lambda u\omega'(u)\, du\right\} d\mu^{(n)}(t), \tag{1.8}$$

where $u^{(n)}$ is the $n$-fold convolution of $\mu$. We define the interval $I_\mu^{(n)} = [na, nb]$ if $b < \infty$ and $I_\mu^{(n)} = [na, \infty)$ if $b = \infty$. Now since $X_1, X_2, \cdots, X_n$ are assumed

to be independent with common distribution function $F(x \mid \lambda)$, we see that the sum $S_n$ is distributed according to (1.8). Furthermore, by applying the factorization criterion for sufficiency to the joint distribution of $X_1, X_2, \cdots, X_n$ (see, e.g., [5]) it is easily seen that $S_n$ is a sufficient statistic for the problem under consideration so that we may confine our attention to decision procedures depending only on the value of $S_n$.

Some particular examples of families of distribution functions $F(x \mid \lambda)$ which are of practical interest are as follows:

*Example* 1: If $\mu$ is the counting measure on the integers zero and one, and $\omega(\lambda) = \ln (\lambda/1 - \lambda)$ for $0 < \lambda < 1$, then $F(x \mid \lambda)$ is the distribution function of a Bernoulli random variable taking on the value one with probability $\lambda$ and zero with probability $1 - \lambda$.

*Example* 2: If $\nu$ is the counting measure on the non-negative integers, $d\mu(x)/d\nu(x) = 1/x!$, and $\omega(\lambda) = \ln \lambda$ for $0 < \lambda < \infty$, then $F(x \mid \lambda)$ is the distribution of a Poisson random variable with expected value $\lambda$.

*Example* 3: If $\nu$ is Lebesgue measure on the positive half-line, $d\mu(x)/d\nu(x) = \eta^{\eta}x^{\eta-1}/\Gamma(\eta)$ for known $\eta > 0$, and $\omega(\lambda) = -\eta/\lambda$, then for each $\eta$, $F(x \mid \lambda)$ is a gamma distribution with $E(X \mid \lambda) = \lambda$ and $\text{Var}(X \mid \lambda) = \lambda^2/\eta$.

In order to discuss the properties of the Bayes sample size it will be necessary to consider a further specialization of the class of distribution functions $F(x \mid \lambda)$. To this end we let

$$\omega(\lambda) = \frac{1}{k} \ln \frac{\lambda}{k\alpha + \beta\lambda}, \tag{1.9}$$

where $k$ is a positive number and $\alpha$ and $\beta$ are numbers such that either (i) $\alpha > 0$ and $\beta \geqq 0$, or (ii) $\alpha > 0$ and $\beta = -\alpha/b^*$ where $b^*$ is a positive integer. Let $\mu(x)$ be a measure such that

$$\frac{d\mu(x)}{d\nu(x)} = \begin{cases} 1, & x = 0, \\ \dfrac{\alpha(\alpha+\beta)\cdots\left(\alpha + \left(\dfrac{x}{k} - 1\right)\beta\right)}{\left(\dfrac{x}{k}\right)!}, & x = k, 2k, \cdots, \end{cases} \tag{1.10}$$

where $\nu$ is counting measure on $0, k, 2k, \cdots$. For case (i) $I_\mu = [0, \infty)$ and for case (ii) $I_\mu = [0, b]$, where $b = kb^*$. These definitions permit us to define the class $\mathfrak{F}_1$ of distribution functions $F(x \mid \lambda)$ as follows:

(1.11) $\mathfrak{F}_1 =$ The class distribution functions $F(x \mid \lambda)$ of the form (1.4) for which the corresponding $\omega(\lambda)$ and $\mu(x)$ are determined by (1.9) and (1.10) respectively.

The class of distribution functions $\mathfrak{F}_1$ clearly contains the Bernoulli ($\alpha = k = 1$, $\beta = -1$) and Poisson ($\alpha = k = 1, \beta = 0$) examples discussed earlier That the distribution functions in the class $\mathfrak{F}_1$ are well defined follows from

THEOREM 2.4: *If the function* $\omega(\lambda)$ *and the measure* $\mu(x)$ *are defined by* (1.9) *and* (1.10), *then condition* (1.3B) *is satisfied.*

The form of the $n$-fold convolution $F^{(n)}(x \mid \lambda)$ of a distribution function $F(x \mid \lambda)$ in the class $\mathfrak{F}_1$ is needed in the derivation of an asymptotic expansion for the Bayes risk. The following theorem gives a formula for $F^{(n)}(x \mid \lambda)$.

THEOREM 2.5: *If* $F(x \mid \lambda) \varepsilon \mathfrak{F}_1$, *then for all integer values of* $m$, $F^{(n)}(x \mid \lambda)$ *is given by*

$$(1.12)\quad F^{(n)}(km \mid \lambda) = \begin{cases} 0, & m \leqq 0, \\ 1 - mr^{(n)}(km) \displaystyle\int_0^\lambda \frac{t^{m-1}}{(k\alpha + \beta t)^m} & \\ \qquad \cdot \exp\left\{-n \displaystyle\int_0^t \frac{\alpha\, du}{(k\alpha + \beta u)}\right\} dt, & m = 1, 2, \cdots, \end{cases}$$

*where*

$$(1.13)\quad r^{(n)}(x) = \begin{cases} 1, & x = 0, \\ \dfrac{n\alpha(n\alpha + \beta) \cdots \left(n\alpha + \left(\dfrac{x}{k} - 1\right)\beta\right)}{\left(\dfrac{x}{k}\right)!}, & x = k, 2k, \cdots. \end{cases}$$

In addition to the class $\mathfrak{F}_1$ of discrete distributions we will consider the class of continuous gamma type families defined by (1.4) with

$$(1.14)\quad \omega(\lambda) = -\eta/\lambda,$$

and

$$\frac{du(x)}{d\nu(x)} = \begin{cases} 0, & x < 0, \\ \dfrac{\eta^\eta x^{\eta-1}}{\Gamma(\eta)}, & x \geqq 0, \end{cases}$$

where $\nu(x)$ is Lebesgue measure on the positive half-line. The class $\mathfrak{F}_2$ is defined by

(1.16) $\mathfrak{F}_2$ = the family of distribution functions $F(x \mid \lambda)$ of the form (1.4) with $\omega(\lambda)$ and $\mu(x)$ given by (1.14) and (1.15).

This definition leads to

THEOREM 2.6: *For any distribution function* $F(x \mid \lambda) \varepsilon \mathfrak{F}_2$, (*i*) *condition* (1.3B) *is satisfied, and* (*ii*)

$$(1.17)\quad F^{(n)}(x \mid \lambda) = \frac{(\eta x)^{n\eta}}{\Gamma(n\eta)} \int_\lambda^\infty u^{-n\eta-1} \exp\left\{-\frac{x\eta}{u}\right\} du.$$

Returning now to the underlying decision problem, for any fixed $n$ let $\delta(s_n)$ be a decision rule which is to be interpreted as the probability of acceptance of the uninspected remainder of the lot when $s_n$ is the observed value of the sufficient statistic $S_n$. Regarding $\lambda$ as a random variable $\Lambda$ by virtue of the assumed

existence of an *a priori* probability distribution we observe that, given the value $\lambda$ of $\Lambda$, $S_N - S_n$ and $S_n$ are conditionally independently distributed according to $F^{(N-n)}(x \mid \lambda)$ and $F^{(n)}(x \mid \lambda)$ respectively so that

$$\begin{aligned} E\{S_N - S_n \mid S_n\} &= E\{E\{S_N - S_n \mid \Lambda, S_n\} S_n\} \\ &= E\{(N - n)\Lambda \mid S_n\} \\ &= (N - n)E\{\Lambda \mid S_n\}. \end{aligned} \tag{1.18}$$

Hence, referring to (1.1) the risk incurred by using the rule $\delta$ may be written as

$$\begin{aligned} R(\delta, n, N) &= E\{\delta(S_n)[a_1(S_N - S_n) + a_2(N - n)]\} \\ &\quad + E\{[1 - \delta(S_n)][r_1(S_N - S_n) + r_2(N - n)]\} \\ &\quad + E\{s_1 S_n + s_2 n\} \\ &= E\{\delta(S_n)[a_1 - r_1)(S_N - S_n) + (a_2 - r_2)(N - n)]\} \\ &\quad + (s_1 n + r_1(N - n))E(\Lambda) + s_2 n + r_2(N - n) \\ &= (N - n)E\{\delta(S_n)[(a_1 - r_1)E(\Lambda \mid S_n) + a_2 - r_2\} \\ &\quad + [s_1 n + r_1(N - n)]E(\Lambda) + s_2 n + r_2(N - n). \end{aligned} \tag{1.19}$$

From this it is clear that the essentially unique Bayes decision rule is given by

$$\delta^*(s_n) = \begin{cases} 1, & E\{\Lambda \mid S_n = s_n\} \leqq c, \\ 0, & \text{otherwise}, \end{cases} \tag{1.20}$$

where $c = r_2 - a_2/a_1 - r_1$, provided $a < c < b$. To avoid trivial cases where acceptance or rejection is determined without sampling we shall assume that $a < c < b$, which implies either (1) $r_1 < a_1$ and $r_2 > a_2$, or (2) $r_1 > a_1$ and $r_2 < a_2$. Referring to (1.1) we see that for any given cost situation where (1) holds we may find a corresponding second situation where (2) holds which becomes identical with the first when the two actions are interchanged. Hence in the sequel we shall assume without loss of generality that (2) holds.

Unfortunately, for many *a priori* distributions and many families $F(x \mid \lambda)$ of interest the quantity $E\{\Lambda \mid S_n = s_n\}$ cannot be expressed explicitly. The following results which are proved in Section 3 give more explicit characterizations of $\delta^*$ for the case where $n$ is large. These results are also needed for the determination of the Bayes sample size.

Let $G(\lambda)$ be the *a priori* distribution function of the parameter $\lambda$, i.e., $G(\lambda) = P\{\Lambda < \lambda\}$. We assume that $G(\lambda)$ assigns probability one to the closed interval $[a, b]$. We assume further that $E(\Lambda)$ is finite and that $G(\lambda)$ does not assign probability one to any single point. Define the function $\varphi_n(t)$ for $t \varepsilon I_\mu^{(n)}$ by

$$\varphi_n(t) = \frac{\int_0^\infty \lambda \exp\left\{t\omega(\lambda) - n\int_\gamma^\lambda u\omega'(u)\,du\right\} dG(\lambda)}{\int_0^\infty \exp\left\{t\omega(\lambda) - n\int_\gamma^\lambda u\omega'(u)\,du\right\} dG(\lambda)}. \tag{1.21}$$

It is easily verified that $\varphi_n(t)$ coincides with $E\{\Lambda \mid S_n = t\}$ for almost all values of $t$ which are possible values of $S_n$. From this definition we obtain

THEOREM 3.1: *The function $\varphi_n(t)$ given by (1.21) is finite and strictly increasing for $t \varepsilon I_\mu^{(n)}$.*

We now observe that exactly one of the following must hold:

$$\text{(1.22)} \qquad \begin{aligned} &\text{i) } \varphi_n(nb) < c,\\ &\text{ii) } \varphi_n(na) > c,\\ &\text{iii) } \varphi_n(t_n) = c \text{ for a unique } t_n \,\varepsilon\, I_\mu^{(n)}. \end{aligned}$$

Hence the Bayes decision rule $\delta^*(S_n)$ given by (1.20) is equivalently expressed by

$$\text{(1.23)} \qquad \delta^*(S_n) = \begin{cases} 1, & \text{if } S_n \leqq t(n),\\ 0, & \text{otherwise,} \end{cases}$$

where

$$\text{(1.24)} \qquad t(n) = \begin{cases} na - 1, & \varphi_n(nb) < c,\\ nb, & \varphi_n(na) > c,\\ t_n, & \varphi_n(t_n) = c,\ t_n \,\varepsilon\, I_\mu^{(n)}. \end{cases}$$

An asymptotic characterization of the function $t(n)$ for *a priori* distributions $G(\lambda)$ placing positive weight on both sides of $c$ is given by

THEOREM 3.2. *If* $\lambda_0 = \sup\{\lambda:\lambda \leqq c; G(\lambda+) - G(\lambda - \epsilon) > 0 \text{ all } \epsilon > 0\}$ *and* $\lambda_1 = \inf\{\lambda:\lambda \geqq c; G(\lambda + \epsilon) - G(\lambda) > 0, \text{ all } \epsilon > 0\}$, *then*

$$\text{(1.25)} \qquad \lambda_0 \leqq \liminf_{n\to\infty} \frac{t(n)}{n} \leqq \limsup_{n\to\infty} \frac{t(n)}{n} \leqq \lambda_1 .$$

Thus for the particular case where $G(\lambda)$ assigns positive weight to every interval about $c$ we have $\lambda_0 = \lambda_1$ and

$$\text{(1.26)} \qquad t(n) = cn + o(n).$$

In order to obtain a more precise asymptotic characterization of $t(n)$ we define two classes of *a priori* distributions as follows:

(1.27) $\mathcal{G}_1$ = the class of all $G(\lambda)$ which are twice continuously differentiable in some open interval about $c$ with $G'(c) > 0$;

(1.28) $\mathcal{G}_2$ = the class of all $G(\lambda)$ for which there exist numbers $l_G$ and $u_G$, $l_G < c < u_G$, which are assigned positive weight by $G(\lambda)$ and are such that $G(u_G) - G(c+) = 0$ and $G(c) - G(l_G+) = 0$.

The class of *a priori* distributions assigning probability one to a finite set of points is of course a subset of $\mathcal{G}_2$. We now have

THEOREM 3.3: *If* $G(\lambda) \,\varepsilon\, \mathcal{G}_1$, *then*

$$\text{(1.29)} \qquad t(n) = cn + \frac{\omega''(c)}{(\omega'(c))^2} - \frac{G''(c)}{G'(c)\omega'(c)} + o(1).$$

THEOREM 3.4: *If* $G(\lambda) \varepsilon \mathcal{G}_2$, $\zeta_1 = G(l_G+) - G(l_G)$, *and* $\zeta_2 = G(u_G+) - G(u_G)$, *then*

$$(1.30) \qquad t(n) = n \frac{\int_{l_G}^{u_G} u\omega'(u)\,du}{\omega(u_G) - \omega(l_G)} + \frac{\ln \dfrac{\zeta_1(c - l_G)}{\zeta_2(u_G - c)}}{\omega(u_G) - \omega(l_G)} + o(1).$$

Although the classes $\mathcal{G}_1$ and $\mathcal{G}_2$ are not exhaustive we have characterized $t(n)$ and hence $\delta^*(s_n)$ sufficiently for most practical purposes as long as $n$ is reasonably large. We now turn our attention to the problem of determining the Bayes sample size $n^* = n^*(N)$ which minimizes the risk $R(\delta^*, n, N)$, and seek an asymptotic characterization of the Bayes sample size $n^*(N)$ for large $N$.

The parameter $k$ appearing in (1.9) is essentially a scale parameter in the sense that if $X$ is a random variable distributed according to (1.9) with $k = 1$ and if $\lambda$ is replaced by $\lambda^*/k$ then $kX$ has a distribution of the same form as (1.9) with arbitrary $k$ and with $\lambda^*$ playing the role of $\lambda$. A similar remark applys to the parameter $\eta$ appearing in (1.15). Hence the cases where $k$ and $\eta$ are arbitrary may be obtained from the cases where $k = 1$ and $\eta = 1$ by multiplying the appropriate cost coefficients by $k$ or $\eta$ and making suitable changes of variables in the *a priori* distribution functions. For the sake of simplicity the remaining results are stated for the cases $k = 1$ and $\eta = 1$ only.

The asymptotic behavior of $n^*(N)$ is characterized by the following theorems, which are proved in Section 4.

THEOREM 4.1: *If* $F(x \mid \lambda) \varepsilon \mathfrak{F}_1$ (*with* $k = 1$) and $G(\lambda) \varepsilon \mathcal{G}_1$, *then the Bayes risk for fixed n and N is given by*

$$(1.31) \qquad \begin{aligned} R(\delta^*, n, N) = {} & n((s_1 - r_1)E(\Lambda) + (s_2 - r_2) + (r_1 - a_1)\int_0^c (\lambda - c)\,dG(\lambda)) \\ & + N(r_1 E(\Lambda) + r_2 + (a_1 - r_1)\int_0^c (\lambda - c)\,dG(\lambda)) \\ & + (N - n)\frac{(r_2 - a_2)(\alpha + \beta c)G'(c)}{2\alpha n} + (N - n)o\left(\frac{1}{n}\right) \end{aligned}$$

*and the Bayes sample size is*

$$(1.32) \qquad n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ N^{1/2}\left(\dfrac{(r_2 - a_2)(\alpha + \beta c)G'(c)}{2\alpha A_G}\right)^{1/2} + o(N^{1/2}), & A_G > 0, \end{cases}$$

*where*

$$(1.33) \qquad A_G = (s_1 - r_1)E(\Lambda) + s_2 - r_2 + (r_1 - a_1)\int_0^c (\lambda - c)\,dG(\lambda).$$

THEOREM 4.2: *If* $F(x \mid \lambda) \varepsilon \mathfrak{F}_1$ (*with* $k = 1$), $G(\lambda) \varepsilon \mathcal{G}_2$ *and* $A_G$ *is defined by* (1.33), *then the Bayes sample size is given by*

$$(1.34)\qquad n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ K \ln N - \dfrac{K}{2} \ln \ln N + O(1), & A_G > 0. \end{cases}$$

(The definition of $K$ is lengthy and is contained in the proof in Section 4.)

THEOREM 4.3: *If $F(x \mid \lambda) \varepsilon \mathfrak{F}_2$ (with $\eta = 1$), $G(\lambda) \varepsilon \mathcal{G}_1$ and $A_G$ is defined by (1.33), then the Bayes sample size is given by*

$$(1.35)\qquad n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ N^{1/2} \left( \dfrac{(r_2 - a_2)cG'(c)}{2A_G} \right)^{1/2} + o(N^{1/2}), & A_G > 0. \end{cases}$$

THEOREM 4.4: *If $F(x \mid \lambda) \varepsilon \mathfrak{F}_2$ (with $\eta = 1$), $G(\lambda) \varepsilon \mathcal{G}_2$ and $A_G$ is defined by (1.33), then the Bayes sample size is given by*

$$(1.36)\qquad n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ K \ln N - \dfrac{K}{2} \ln \ln N + O(1), & A_G > 0. \end{cases}$$

In all cases where $A_G \leqq 0$ the proper procedure is to screen the lot completely (i.e., take $n^*(N) = N$). Theorems 4.1 and 4.3 show that if an *a priori* probability density for the parameter $\lambda$ exists in the vicinity of the critical point $c$ and if this density is smooth and positive at $c$, then the optimal sample size when $A_G > 0$ is approximately proportional to the square root of the lot size $N$ when $N$ is large. For the cases covered by Theorems 4.2 and 4.4 where the *a priori* probability that $\lambda$ lies within a certain neighborhood of $c$ is zero, the optimal sample size when $A_G > 0$ is approximately proportional to the logarithm of $N$ when $N$ is large. It is clear from these results that the optimal rate of increase for the sample size depends critically on the fine structure of the *a priori* information about $\lambda$ in the vicinity of $c$. This is especially remarkable in view of the fact that $c$ is actually the "indifference" value of $\lambda$ in the sense that if $\lambda = c$ then either acceptance or rejection of the lot leads to the same expected cost.

Referring to (1.3) and (1.32) of Theorem 4.1 we may write

$$(1.37)\qquad R(\delta^*, n, N) = A_G n + B_G N + C_G(N - n) \left( \frac{1}{n} + o\left(\frac{1}{n}\right) \right),$$

and if $A_G > 0$

$$(1.38)\qquad n^*(N) = \left( \frac{C_G}{A_G} \right)^{1/2} N^{1/2} + o(N^{1/2}),$$

where $A_G$ is given by (1.33), and $B_G$ and $C_G$ are coefficients depending on the costs and the *a priori* distribution $G$. It is easily verified that the term $B_G N$ represents the "subminimal" risk which would result if the value $\lambda$ of $\Lambda$ were known exactly without sampling and the decision to accept or reject determined ac-

cordingly. From (1.37) and (1.38) we obtain

$$R(\delta^*, n^*(N), N) = B_G N + 2(A_G C_G)^{1/2} N^{1/2} + o(N^{1/2}), \tag{1.39}$$

which shows that the amount by which the Bayes risk exceeds the subminimal risk due to the uncertainty concerning the value of $\Lambda$ is of smaller order in $N$ than the subminimal risk itself. Expression (1.39) is still valid if the sample size is determined by taking only the first term in the asymptotic expansion for $n^*(N)$ so that not much is lost by making this approximation if $N$ is large. Similar remarks may be made for the cases covered by Theorems 4.2, 4.3 and 4.4. For the cases covered by Theorems 4.2 and 4.4 the term added to the subminimal risk in the expressions for $R(\delta^*, n^*(N), N)$ is of order $\ln N$.

## 2. Theorems concerning the class of distribution functions $F(x \mid \lambda)$.

THEOREM 2.1: *The function $\omega(\lambda)$ given by (1.9) is unique and $d\omega(\lambda)/d\lambda$ exists and is positive for $\lambda \varepsilon I_\mu$.*

PROOF: By assumption (1.3) there exist numbers $\omega_0$, $\omega_1$ such that the ratio

$$\rho(\omega) = \frac{\int_{[a,\infty)} xe^{\omega x}\, d\mu(x)}{\int_{[a,\infty)} e^{\omega x}\, d\mu(x)} \tag{2.1}$$

is finite for $\omega_0 < \omega < \omega_1$. Furthermore, $\rho(\omega)$ is differentiable with respect to $\omega$ and

$$\rho'(\omega) = \int_{[a,\infty)} (x - \rho(\omega))^2 \frac{e^{\omega x}}{\int_{[a,\infty)} e^{\omega x}\, d\mu(x)}\, d\mu(x) > 0, \tag{2.2}$$

for $\omega_0 < \omega < \omega_1$. Now for $\lambda \varepsilon I_\mu$ we have, by (1.3B), $\rho(\omega(\lambda)) = \lambda$ which implies that $\omega(\lambda)$ is unique and that $d\omega(\lambda)/d\lambda$ exists and is given by

$$\frac{d\omega(\lambda)}{d\lambda} = \frac{1}{\rho'(\omega(\lambda))} > 0. \tag{2.3}$$

THEOREM 2.2: *If $F(x \mid \lambda)$ is defined by (1.4), then all moments of $F(x \mid \lambda)$ exist, and all derivatives of $\omega(\lambda)$ exist and are finite for $\lambda \varepsilon I_\mu$.*

PROOF: The function $\omega(\lambda)$ is continuous and strictly increasing for $\lambda \varepsilon I_\mu$ by Theorem 2.1. Therefore, the moment generating function given by

$$m_\lambda(t) = \frac{\int_{[a,\infty)} e^{(t+\omega(\lambda))x}\, d\mu(x)}{\int_{[a,\infty)} e^{\omega(\lambda)x}\, d\mu(x)} \tag{2.4}$$

exists for each $\lambda \varepsilon I_\mu$ for all $t$ in some open interval about zero since both the numerator and the denominator of the ratio (1.3B) must be finite for any $\lambda \varepsilon I_\mu$. That is, for any fixed $\lambda \varepsilon I_\mu$, we may choose $t \neq 0$ small enough in magni-

tude so that there exists a $\lambda^* \varepsilon I_\mu$ for which $|t| + \omega(\lambda) < \omega(\lambda^*)$ so that the integral in the numerator must converge.

Repeated formal differentiations of $\rho(\omega)$ yield sums of ratios involving products of integrals of the form $\int_{[a,\infty)} x^k e^{\omega x} \, d\mu(x)$ in the numerators and powers of $\int_{[a,\infty)} e^{\omega x} \, d\mu(x)$ in the denominators. These integrals are finite and those in the denominators do not vanish for $\omega = \omega(\lambda)$, $\lambda \varepsilon I_\mu$, so all derivatives of $\rho(\omega)$ exist for such values of $\omega$. As before, for $\lambda \varepsilon I_\mu$,

$$\frac{d\omega(\lambda)}{d\lambda} = \frac{1}{\rho'(\omega(\lambda))} \tag{2.5}$$

and repeated application of the rule for differentiation of implicit functions shows that $\omega(\lambda)$ possesses derivatives of all orders for $\lambda \varepsilon I_\mu$.

THEOREM 2.3: *The distribution function $F(x \mid \lambda)$ defined by (1.4) may be represented for $a < x \leqq b$ and for $\lambda \varepsilon I_\mu$ by*

$$F(x \mid \lambda) = K(\gamma) \int_{[a,x)} \exp\left\{\omega(\lambda) - \int_\gamma^\lambda u\omega'(u) \, du\right\} d\mu(t), \tag{2.6}$$

*if and only if assumption (1.3B) is satisfied, where $\gamma \varepsilon I_\mu$ and $K(\gamma)$ is a normalizing factor depending on the choice of $\gamma$ determined so that $F(b+ \mid \lambda) = 1$.*

PROOF: We observe that

$$\frac{d}{d\lambda} \int_{[a,\infty)} e^{\omega(\lambda)x} \, d\mu(x) = \omega'(\lambda) \int_{[a,\infty)} x e^{\omega(\lambda)x} \, d\mu(x) \tag{2.7}$$

so that dividing both sides by $\int_{[a,\infty)} e^{\omega(\lambda)x} \, d\mu(x)$ and referring to assumption (1.3B) we have for $\lambda \varepsilon I_\mu$,

$$\frac{d}{d\lambda} \ln \int_{[a,\infty)} e^{\omega(\lambda)x} \, d\mu(x) = \lambda\omega'(\lambda). \tag{2.8}$$

Hence

$$\begin{aligned} \int_{[a,\infty)} e^{\omega(\lambda)x} \, d\mu(x) &= \exp\left\{\int \lambda\omega'(\lambda) \, d\lambda + c\right\} \\ &= K(\gamma) \exp\left\{\int_\gamma^\lambda u\omega'(u) \, du\right\} \end{aligned} \tag{2.9}$$

for $\gamma \varepsilon I_\mu$. The fact that assumption (1.3B) is satisfied whenever (2.6) is valid follows immediately by differentiating the expression $F(b+ \mid \lambda) = 1$ with respect to $\lambda$.

We now verify that the classes $\mathfrak{F}_1$ and $\mathfrak{F}_2$ defined by (1.11) and (1.16) satisfy the above assumptions, and determine closed expressions for the $n$-fold convolution of distribution functions in these classes.

THEOREM 2.4: *If the function $\omega(\lambda)$ and the measure $\mu(x)$ are defined by (1.9) and (1.10), then condition (1.3B) is satisfied.*

PROOF: Let the function $r(x)$, $x = 0, k, 2k, \cdots$ be defined by the generating

function

$$(2.10)\qquad \sum_{x=0}^{b^*} r(kx)t^x = \begin{cases} e^{at}, & \beta = 0, \\ (1-\beta t)^{-\alpha/\beta}, & \beta \neq 0. \end{cases}$$

It is easily verified by successive differentiation of (2.10) that $r(x) = d\mu(x)/d\nu(x)$ as given by 1.10. Substituting $\lambda/(k\alpha + \beta\lambda)$ for $t$ in (2.10), we obtain

$$(2.11)\qquad \sum_{x=0}^{b^*} r(kx)\left(\frac{\lambda}{k\alpha+\beta\lambda}\right)^x = \begin{cases} e^{\lambda/k}, & \beta = 0, \\ \left(1+\dfrac{\beta\lambda}{k\alpha}\right)^{\alpha/\beta}, & \beta \neq 0. \end{cases}$$

Noting that $\omega(\lambda) = (1/k)\ln(\lambda/(k\alpha+\beta\lambda))$ by (1.9) we may write (2.11) as

$$(2.12)\qquad \int_{[0,\infty)} e^{x\omega(\lambda)}\,d\mu(x) = \begin{cases} e^{\lambda/k}, & \beta = 0, \\ \left(1+\dfrac{\beta\lambda}{k\alpha}\right)^{\alpha/\beta}, & \beta \neq 0. \end{cases}$$

Differentiating (2.12) with respect to $\lambda$ and dividing both members by $\omega'(\lambda) = (\alpha/\lambda(k\alpha+\beta\lambda))$, we obtain

$$(2.13)\qquad \int_{[0,\infty)} xe^{x\omega(\lambda)}\,d\mu(x) = \begin{cases} \lambda e^{-\lambda/k}, & \beta = 0, \\ \lambda\left(1+\dfrac{\beta\lambda}{k\alpha}\right)^{\alpha/\beta}, & \beta \neq 0. \end{cases}$$

The proof is completed by noting that the ratio of (2.13) to (2.12) is always $\lambda$ so that (1.3B) is satisfied.

Theorems 2.3 and 2.4 imply that for any $F(x \mid \lambda)\ \varepsilon\ \mathfrak{F}_1$

$$(2.14)\qquad F(kx \mid \lambda) = \sum_{t=0}^{x-1} r(kt)\exp\left\{kt\omega(\lambda) - \int_0^\lambda u\omega'(u)\,du\right\},$$

for integer values of $x$, where $\omega(\lambda)$ is given by (1.9), $r(x) = d\mu(x)/d\nu(x)$ is given by (1.10), and the value of $\gamma$ appearing in (2.6) is taken to be zero. The fact that we must have $K(0) = 1$ if $F(b+ \mid \lambda)$ is to equal one follows from the observation that $\omega(\lambda) \to -\infty$ as $\lambda \to 0$ together with the assumption that $r(0) = 1$.

The following theorem provides an integral formula for the $n$-fold convolution $F^{(n)}$ of $F(x \mid \lambda)$ when $F$ is in $\mathfrak{F}_1$.

THEOREM 2.5: *If $F(x \mid \lambda)\ \varepsilon\ \mathfrak{F}_1$, then for integer values of $m$, $F^{(n)}$ is given by*

$$(2.15)\qquad F^{(n)}(km \mid \lambda) = \begin{cases} 0, & m \leq 0, \\ 1 - mr^{(n)}(km)\displaystyle\int_0^\lambda \frac{t^{m-1}}{(k\alpha+\beta t)^m} \cdot\exp\left\{-n\int_0^t \frac{\alpha\,du}{k\alpha+\beta u}\right\}dt, & m = 1, 2, \cdots, \end{cases}$$

*where*

$$(2.16)\qquad r^{(n)}(x) = \begin{cases} 1, & x = 0, \\ \dfrac{n\alpha(n\alpha+\beta)\cdots\left(n\alpha + \left(\dfrac{x}{k} - 1\right)\beta\right)}{\left(\dfrac{x}{k}\right)!}, & x = k, 2k, \cdots. \end{cases}$$

PROOF: By (1.7)

$$(2.17)\qquad F^{(n)}(\infty \mid \lambda) = \int_{[0,\infty)} \exp\left\{\omega(\lambda)x - n\int_0^\lambda u\omega'(u)\,du\right\} d\mu^{(n)}(x) = 1$$

where $\mu^{(n)}$ is the $n$-fold convolution of $\mu$. Hence, letting $r^{(n)}(x) = d\mu^{(n)}(x)/d\nu(x)$ and $t = \lambda/(k\alpha + \beta\lambda)$ and recalling that $\omega(\lambda) = (1/k)\ln(\lambda/(k\alpha + \beta t))$ for this case, we have

$$(2.18)\qquad \begin{aligned} \sum_{x=0}^{n\delta^*} r^{(n)}(kx)t^x &= \exp\left\{n\int_0^{k\alpha t/1-\beta t} \frac{\alpha\,du}{k\alpha + \beta u}\right\} \\ &= \begin{cases} e^{n\alpha t}, & \beta = 0, \\ (1-\beta t)^{-n\alpha/\beta}, & \beta \neq 0. \end{cases} \end{aligned}$$

Successive differentiation of (2.17) with respect to $t$ yields (2.16). Referring to (1.8), we may write

$$(2.19)\qquad F^{(n)}(km \mid \lambda) = \sum_{x=0}^{m-1} r^{(n)}(kx)\left(\frac{\lambda}{k\alpha + \beta\lambda}\right)^x \exp\left\{-n\int_0^\lambda \frac{\alpha\,du}{k\alpha + \beta u}\right\},$$

for integer values of $m$. Assuming that (2.15) holds for some integer $m$, we have

$$(2.20)\qquad \begin{aligned} &F^{(n)}(k(m+1) \mid \lambda) \\ &= F^{(n)}(km \mid \lambda) + r^{(n)}(km)\left(\frac{\lambda}{k\alpha + \beta\lambda}\right)^m \exp\left\{-n\int_0^\lambda \frac{\alpha\,du}{k\alpha + \beta u}\right\} \\ &= 1 - mr^{(n)}(km)\int_0^\lambda \frac{t^{m-1}}{(k\alpha + \beta t)^m}\exp\left\{-n\int_0^t \frac{\alpha\,du}{k\alpha + \beta u}\right\} dt \\ &\qquad + r^{(n)}(km)\left(\frac{\lambda}{k\alpha + \beta\lambda}\right)^m \exp\left\{-n\int_0^\lambda \frac{\alpha\,du}{k\alpha + \beta u}\right\}. \end{aligned}$$

Integrating the second term on the right by parts yields

$$(2.21)\qquad \begin{aligned} &mr^{(n)}(km)\int_0^\lambda t^{m-1}\left[(k\alpha + \beta t)^{-m}\exp\left\{-n\int_0^t \frac{\alpha\,du}{k\alpha + \beta u}\right\}\right] dt \\ &= r^{(n)}(km)(n\alpha + m\beta)\int_0^\lambda \frac{t^m}{(k\alpha + \beta t)^{m+1}}\exp\left\{-n\int_0^t \frac{\alpha\,du}{k\alpha + \beta u}\right\} dt \\ &\qquad + r^{(n)}(km)\left(\frac{\lambda}{k\alpha + \beta\lambda}\right)^m \exp\left\{-n\int_0^\lambda \frac{\alpha\,du}{k\alpha + \beta u}\right\}. \end{aligned}$$

Substituting this expression in (2.20) and recalling (2.16), we see that (2.15) is verified for $F^{(n)}(k(m+1))$. It is easily shown directly that (2.15) holds for $m = 1$ so that the desired result follows by induction.

We now consider the family $\mathfrak{F}_2$ of distribution functions obtained when $\omega(\lambda)$ and $\mu(x)$ are defined by (1.14) and (1.15).

THEOREM 2.6: *If $F(x \mid \lambda) \varepsilon \mathfrak{F}_2$, then (i) condition (1.3B) is satisfied, and (ii)*

$$F^{(n)}(x \mid \lambda) = \frac{(\eta x)^{n\eta}}{\Gamma(n\eta)} \int_\lambda^\infty u^{-n\eta-1} e^{-x\eta/u}\, du. \tag{2.21}$$

PROOF: Any distribution in $\mathfrak{F}_2$ is a gamma distribution with parameters determined so that the first moment is $\lambda$, thereby satisfying (1.3B). The convolution of such gamma distributions is well known (cf. [6]) to be given by

$$F^{(n)}(x \mid \lambda) = \int_0^x \frac{\eta^{n\eta} t^{n\eta-1}}{\Gamma(n\eta)\lambda^{n\eta}} e^{-t\eta/\lambda}\, dt. \tag{2.22}$$

By making the change of variable $u = x\lambda/t$ we obtain (2.21).

**3. The Bayes decision rule.**

THEOREM 3.1: *The function $\varphi_n(t)$ given by (1.21) is finite and strictly increasing for $t \varepsilon I_\mu^{(n)}$.*

PROOF: The finiteness of $E(\Lambda) = \int_0^\infty \lambda\, dG(\lambda)$ insures the finiteness of $\varphi_n(t)$ for all $t \varepsilon I_\mu^{(n)}$ with the possible exception of a set of $\mu^{(n)}$-measure zero since

$$\varphi_n(t) = E\{\Lambda \mid S_n = t\}, \tag{3.1}$$

for all $t \varepsilon I_\mu^{(n)} - A$, where $A$ is the exceptional null set. Hence, for any fixed $t \varepsilon I_\mu^{(n)}$, we may choose $t_1$, $t_2 \varepsilon I_\mu^{(n)}$ such that $t_1 \leqq t \leqq t_2$ and $\varphi_n(t_1)$, $\varphi_n(t_2)$ are finite. Then $t\omega(\lambda) \leqq \max\,(t_1\omega(\lambda), t_2\omega(\lambda))$ for all $\lambda \varepsilon I_\mu$ and the finiteness of $\varphi_n(t)$ follows from the finiteness of the integrals in the expressions for $\varphi_n(t_1)$ and $\varphi_n(t_2)$. Now choose $t$ and $\delta > 0$ so that $[t, t+\delta] \subset I_\mu^{(n)}$ and let

$$H_t(z) = \frac{\displaystyle\int_0^z \exp\left\{t\omega(\lambda) - n\int_\gamma^\lambda u\omega'(u)\, du\right\} dG(\lambda)}{\displaystyle\int_0^\infty \exp\left\{t\omega(\lambda) - n\int_\gamma^\lambda u\omega'(u)\, du\right\} dG(\lambda)}. \tag{3.2}$$

Then $H_t(z)$ may be interpreted as the distribution function of some random variable $Z$ and we may write

$$\varphi_n(t+\delta) - \varphi_n(t) = \frac{E\{Z \exp\{\delta\omega(Z)\}\} - E\{Z\}E\{\exp\{\delta\omega(Z)\}\}}{E\{\exp\{\delta\omega(Z)\}\}}. \tag{3.3}$$

It is intuitively clear and follows rigorously from the inequality on page 43 of [7] that the right hand side of (3.3) is strictly positive for all $\delta > 0$ whenever $G(\lambda)$ and hence $H_t(z)$ are non-degenerate. Thus $\varphi_n(t)$ is strictly increasing and the proof is completed.

We now recall that the Bayes decision rule (1.20) is equialvent to

$$\delta^*(S_n) = \begin{cases} 1, & S_n \leqq t(n), \\ 0, & S_n > t(n), \end{cases} \tag{3.4}$$

where $t(n)$ is defined by (1.24).

THEOREM 3.2: *If* $\lambda_0 = \sup \{\lambda : \lambda \leqq c; G(\lambda+) - G(\lambda - \epsilon) > 0,$ *all* $\epsilon > 0\}$ *and* $\lambda_1 = \inf \{\lambda : \lambda \geqq c; G(\lambda + \epsilon) - G(\lambda) > 0,$ *all* $\epsilon > 0\}$, *and if* $t(n)$ *is defined by* (1.24), *then*

$$\lambda_0 \leqq \liminf_{n\to\infty} \frac{t(n)}{n} \leqq \limsup_{n\to\infty} \frac{t(n)}{n} \leqq \lambda_1 . \tag{3.5}$$

PROOF: Let $I$ be an indicator function defined on the product of the space of all sequences of numbers $\{s_n\}$ and the real line by

$$I(\{s_n\}, \lambda) = \begin{cases} 1, & \text{if } s_n/n \to \lambda \quad \text{as} \quad n \to \infty, \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

Given $\Lambda = \lambda$, $S_n$ is a sum of conditionally independent identically distributed random variables with mean $\lambda$. Hence, by the strong law of large numbers

$$E\{I(\{S_n\}, \lambda) \mid \Lambda = \lambda\} = 1, \qquad \text{a.s.,} \tag{3.7}$$

and this can be shown to be equivalent to

$$E\{I(\{S_n\}, \Lambda) \mid \Lambda = \lambda\} = 1, \qquad \text{a.s.} \tag{3.8}$$

Thus

$$P\{S_n/n \to \Lambda\} = E\{E\{I(\{S_n\}, \Lambda\} \mid \Lambda\}\} = 1. \tag{3.9}$$

Furthermore, by a martingale convergence theorem (cf., p. 398 of [8]),

$$P\{E(\Lambda \mid S_n) \to \Lambda\} = 1. \tag{3.10}$$

Therefore for all $x$ which are continuity points of $G(x)$,

$$P\{E(\Lambda \mid S_n) < x\} \to G(x), \tag{3.11}$$

and

$$P\{S_n/n < x\} \to G(x), \tag{3.12}$$

as $n \to \infty$. Suppose that $\limsup_{n\to\infty} (t(n)/n) > \lambda_1$. Then there is a $\delta > 0$ such that $\lambda_1 + \delta$ is a continuity point of $G(x)$, and $t(n)/n > \lambda_1 + \delta$ for arbitrarily large values of $n$. Hence, for these values of $n$

$$P\{E(\Lambda \mid S_n) \leqq c\} = P\{S_n/n \leqq t(n)/n\} \geqq P\{S_n/n \leqq \lambda_1 + \delta\}. \tag{3.13}$$

However, as $n \to \infty$

$$P\{S_n/n < \lambda_1 + \delta\} \to G(\lambda_1 + \delta) > G(\lambda_1+) \tag{3.14}$$

and for sufficiently large $n$

$$P\{E(\Lambda \mid S_n) \leqq c\} \leqq G(\lambda_1+) + \epsilon \tag{3.15}$$

for arbitrary $\epsilon > 0$. If we choose $\epsilon > 0$ such that $\epsilon < G(\lambda_1 + \delta) - G(\lambda_1+)$ we are led to a contradiction and hence, $\limsup_{n\to\infty} t(n)/n \leqq \lambda_1$. A similar argument establishes the other inequality of the theorem.

THEOREM 3.3: *If* $G(\lambda) \varepsilon \mathcal{G}_1$, *then*

$$t(n) = cn + \frac{\omega''(c)}{(\omega'(c))^2} - \frac{G''(c)}{G'(c)\omega'(c)} + o(1). \tag{3.16}$$

The proof of Theorem 3.3 will require a sequence of preliminary lemmas, the first two of which will also be used in the later derivation of an asymptotic expansion for the Bayes risk.

Since $P\{E(\Lambda \mid S_n) \to \Lambda\} = 1$, there exists a $\delta > 0$ such that (i) $c + \delta$ is a continuity point of $G(x)$, (ii) $G(c + \delta) < 1$, and (iii) as $n \to \infty$

$$P\{E(\Lambda \mid S_n) > c + \delta\} \to 1 - G(c + \delta) > 0. \tag{3.17}$$

But $E(\Lambda \mid S_n) \leqq \varphi_n(nb)$, a.s., hence $\varphi_n(nb) > c$ for all sufficiently large $n$. Similarly, $\varphi_n(na) < c$ for all sufficiently large $n$. Hence, referring to (1.22) we see that for all sufficiently large $n$, $t(n)$, defined by (1.24), is the solution to

$$I(n) = \int_0^\infty (\lambda - c) \exp\left\{t(n)\omega(\lambda) - n \int_\gamma^\lambda u\omega'(u)\, du\right\} dG(\lambda) = 0. \tag{3.18}$$

The result of Theorem 3.2 suggests that if $G(\lambda) \varepsilon \mathcal{G}_1$, then we should write $t(n) = cn + \psi(n)$ so that

$$I(n) = \int_0^\infty (\lambda - c) \exp\{nh(\lambda) + \psi(n)\omega(\lambda)\}\, dG(\lambda) \tag{3.19}$$

where

$$h(\lambda) = c\omega(\lambda) - \int_\gamma^\lambda u\omega'(u)\, du. \tag{3.20}$$

This form of $I(n)$ will be convenient for the application of results from the theory of the asymptotic expansion of integrals.

LEMMA 3.1: *Let $g(t)$ be any function integrable with respect to a distribution function $H(t)$, and let $\varphi(n, t)$ be a function, $\{t_n\}$ a sequence and $c$ a number such that for all sufficiently small $\epsilon > 0$ there exists a $\delta > 0$ such that for all sufficiently large $n$, $\varphi(n, t) < \varphi(n, t_n) - \delta$ whenever $|t - c| \geqq \epsilon$. Then for any fixed $m$*

$$\int_{-\infty}^{c-\epsilon} g(t) \exp\{n\varphi(n, t)\}\, dH(t) = o(\exp\{n\varphi(n, t_n)\}n^{-m}), \tag{3.21}$$

*and*

$$\int_{c+\epsilon}^{\infty} g(t) \exp\{n\varphi(n, t)\}\, dH(t) = o(\exp\{n\varphi(n, t_n)\}n^{-m}). \tag{3.22}$$

PROOF: For all sufficiently large $n$,

$$\left| \int_{-\infty}^{c-\epsilon} g(t) \exp\{n\varphi(n, t)\} \, dH(t) \right| \leq \exp\{n\varphi(n, t_n) - n\delta\} \int_{-\infty}^{c-\epsilon} |g(t)| \, dH(t), \tag{3.23}$$

and (3.21) follows immediately. Expression (3.22) follows by the same argument.

LEMMA 3.2: *If $g(t)$ is any function which is four times continuously differentiable in some interval containing $c$, and if $g'(c) = 0, g''(c) < 0$, then*

$$\begin{aligned}\int_{c-\epsilon}^{c+\epsilon} (t-c)^r \exp\{ng(t)\} \, dt = \exp\{ng(c)\} \Bigg\{ & n^{-\frac{r+1}{2}} \left(-\frac{2}{g''(c)}\right)^{\frac{r+1}{2}} \\ \cdot \left(\frac{1 + (-1)^r}{2}\right) \Gamma\left(\frac{r+1}{2}\right) & + n^{-\frac{r+2}{2}} \left(\frac{g'''(c)}{3!}\right)\left(-\frac{2}{g''(c)}\right)^{\frac{r+4}{2}} \\ \cdot \Gamma\left(\frac{r+4}{2}\right)\left(\frac{1 + (-1)^{r+1}}{2}\right) & + n^{-\frac{r+3}{2}} \Bigg[ \left(\frac{1 + (-1)^r}{2}\right)\left(-\frac{2}{g''(c)}\right)^{\frac{r+5}{2}} \\ \cdot \Bigg\{ \left(\frac{g''''(c)}{4!}\right) \Gamma\left(\frac{r+5}{2}\right) - & \frac{\left(\frac{g'''(c)}{3!}\right)^2 \Gamma\left(\frac{r+7}{2}\right)}{g''(c)} \Bigg\} \Bigg] + o\left(n^{-\frac{r+3}{2}}\right) \Bigg\}.\end{aligned} \tag{3.24}$$

This is a standard result from the theory of asymptotic expansions and will not be proved here. A proof is outlined, for example, in [9].

The next lemma establishes the boundedness of $\psi(n)$.

LEMMA 3.3: *If $G(\lambda) \varepsilon \mathcal{G}_1$ and $t(n) = cn + \psi(n)$, then $\psi(n) = O(1)$.*

PROOF: The method of proof is to derive an asymptotic expansion for $I(n)$ as defined by (3.19) and show that the assumption that $|\psi(n)| \to \infty$ as $n \to \infty$ leads to a contradiction.

The expression $h(\lambda) + (\psi(n)/n)\omega(\lambda)$ is maximized when $\lambda = c + (\psi(n)/n)$ and, noting that $\psi(n)/n \to 0$ as $n \to \infty$ by Theorem 3.2, it is easily verified that $h(\lambda) + (\psi(n)/n)\omega(\lambda)$ has the properties of the function $\varphi(n, t)$ of Lemma 3.1 with $t = \lambda$ and $t_n = c + (\psi(n)/n)$. Hence choosing $\epsilon > 0$ such that $dG(\lambda) = G'(\lambda) \, d\lambda$ for $\lambda$ in $(c - \epsilon, c + \epsilon)$ we have from (3.19) by Lemma 3.1

$$\begin{aligned} I(n) = & \int_{c-\epsilon}^{c+\epsilon} (\lambda - c) \exp\{nh(\lambda) + \psi(n)\omega(\lambda)\} G'(\lambda) \, d\lambda \\ & + o\left(\exp\left\{nh\left(c + \frac{\psi(n)}{n}\right) + \psi(n)\omega\left(c + \frac{\psi(n)}{n}\right)\right\} n^{-m}\right), \end{aligned} \tag{3.25}$$

for any $m \geq 0$. By the definition of $\mathcal{G}_1$, $G'(\lambda) = G'(c) + O(\lambda - c)$ for $c - \epsilon < \lambda < c + \epsilon$. Hence

$$I(n) = G'(c)\int_{c-\epsilon}^{c+\epsilon} (\lambda - c)\exp\{nh(\lambda) + \psi(n)\omega(\lambda)\}\,d\lambda$$

$$(3.26)\qquad + O\left(\int_{c-\epsilon}^{c+\epsilon} (\lambda - c)^2 \exp\{nh(\lambda) + \psi(n)\omega(\lambda)\}\,d\lambda\right)$$

$$+ o\left(\exp\left\{nh\left(c + \frac{\psi(n)}{n}\right) + \psi(n)\omega\left(c + \frac{\psi(n)}{n}\right)\right\} n^{-m}\right).$$

Letting $\tau_n = \psi(n)/n$ and $t = \lambda - c - \tau_n$, and expanding the exponent in the integrands about $t = 0$ we obtain

$$I(n) = \exp\{nh(c + \tau_n) + \psi(n)\omega(c + \tau_n)\}\Big[G'(c)\int_{-\epsilon-\tau_n}^{\epsilon-\tau_n} (t + \tau_n)$$

$$\cdot\exp\{t^2[nh''(c + \tau_n) + \psi(n)\omega''(c + \tau_n)] + \xi(n, t)\}\,dt\Big]$$

$$(3.27)\qquad + O\Big(\int_{-\epsilon-\tau_n}^{\epsilon-\tau_n} (t + \tau_n)^2 \exp\{t^2[nh''(c + \tau_n) + \psi(n)\omega''(c + \tau_n)]$$

$$+ \xi(n, t)\}\,dt\Big) + o(\exp\{nh(c + \tau_n) + \psi(n)\omega(c + \tau_n)\}n^{-m})$$

where $|\xi(n, t)| \leq knt^3$ for some $k > 0$, all $n$, and all $t$ in $(-\epsilon - \tau_n, \epsilon - \tau_n)$.

Since $\tau_n \to 0$, changing the range of integration from $(-\epsilon - \tau_n, \epsilon - \tau_n)$ to $(-\epsilon^*, \epsilon^*)$ for $0 < \epsilon^* < \epsilon$ adds only terms of negligibly small order by Lemma 3.1. Hence, $I(n)$ may be expressed in terms of integrals of the form

$$(3.28)\qquad \int_{-\epsilon^*}^{\epsilon^*} t^r \exp\{t^2[nh''(c + \tau_n) + \psi(n)\omega''(c + \tau_n)] + \xi(n, t)\}\,dt$$

for $r = 0, 1, 2$. Applying Lemma 3.2 to these integrals, regarding

$$[nh''(c + \tau_n) + \psi(n)\omega''(c + \tau_n)]$$

as the parameter which becomes large, and noting that the first terms in the expansions remain unchanged if either the upper or lower bound for $\xi(n, t)$ is used, we obtain

$$I(n) = \exp\{nh(c + \tau_n) + \psi(n)\omega(c + \tau_n)\}$$

$$(3.29)\qquad \cdot\Big[G'(c)\frac{\psi(n)}{n^{3/2}}\left(\frac{-\pi}{h''(c + \tau_n) + \tau_n\,\omega''(c + \tau_n)}\right)^{1/2}$$

$$+ o\left(\frac{\psi(n)}{n^{3/2}}\right) + O\left(\frac{1}{n^{3/2}}\right)\Big].$$

However, if $|\psi(n_k)| \to \infty$ for any subsequence $n_k \to \infty$, (3.29) implies that $I(n) \neq 0$ for arbitrarily large values of $n$. This, however, is a contradiction, hence, $\psi(n) = O(1)$.

These lemmas now permit us to complete the proof of the theorem.

PROOF OF THEOREM 3.3: Expanding

$$f(n, \lambda) = G'(\lambda) \exp \psi(n)\omega(\lambda) \tag{3.30}$$

about $\lambda = c$ in (3.25) and regarding $h(\lambda)$ as the function $\varphi(n, t)$ appearing in Lemma 3.1 with $t = \lambda$ and $t_n \equiv c$, we have

$$\begin{aligned} I(n) = \int_{c-\epsilon}^{c+\epsilon} (\lambda - c) \exp \{nh(\lambda) + \psi(n)\omega(c)\}[G'(c) \\ + (\lambda - c)(G''(c) + G'(c)\omega(c)\psi(n)) \\ + R(n, \lambda)] \, d\lambda + o(\exp \{nh(c)\}n^{-m}) \end{aligned} \tag{3.31}$$

for any $m \geqq 0$, where

$$R(n, \lambda) = (\lambda - c) \left[\frac{\partial}{\partial\lambda} f(n, c + \theta(n, \lambda)(\lambda - c)) - \frac{\partial}{\partial\lambda} f(n, c)\right] \tag{3.32}$$

for some $0 < \theta(n, \lambda) < 1$. Now let

$$T(n) = \int_{c-\epsilon}^{c+\epsilon} (\lambda - c) \exp \{nh(\lambda)\}R(n, \lambda) \, d\lambda. \tag{3.33}$$

For any arbitrarily small $\delta > 0$ we can find an $\eta(\delta) > 0$ such that $| R(n, \lambda) | < \delta(\lambda - c)$ for $\lambda$ in $(c - \eta(\delta), c + \eta(\delta))$, since

$$\partial/\partial\lambda f(n, \lambda) = [G''(\lambda) + G'(\lambda)\psi(n)\omega'(\lambda)] \exp \{\psi(n)\omega(\lambda)\}$$

regarded as a function of $\lambda$ is continuous at $c$ uniformly in $n$. Applying Lemma 3.2 we have

$$\begin{aligned} | T(n) | &= \left| \int_{c-\eta(\delta)}^{c+\eta(\delta)} R(n,\lambda)(\lambda - c) \exp \{nh(\lambda)\} \, d\lambda \right| + o(\exp \{nh(c)\}n^{-m}) \\ &< \delta \int_{c-\eta(\delta)}^{c+\eta(\delta)} (\lambda - c)^2 \exp \{nh(\lambda)\} \, d\lambda + o(\exp \{nh(c)\}n^{-m}) \\ &= \delta O\left(\frac{\exp \{nh(c)\}}{n^{3/2}}\right). \end{aligned} \tag{3.34}$$

But $\delta$ may be taken arbitrarily small, hence

$$\limsup_{n\to\infty} (| T(n) |/\exp \{nh(c)\}n^{-3/2})$$

is less than an arbitrarily small quantity, so that $T(n) = o(\exp \{nh(c)\}n^{-3/2})$. Using this fact and applying Lemma 3.2 to the terms of (3.31) involving $(\lambda - c)$ and $(\lambda - c)^2$ we have,

$$I(n) = n^{-3/2} \exp\{nh(c) + \psi(n)\omega(c)\}\left[G'(c)\left(\frac{h'''(c)}{3!}\right)\right.$$

$$(3.35) \qquad \cdot\left(-\frac{2}{h''(c)}\right)^{5/2} \Gamma(5/2) + (G''(c) + \psi(n)\omega'(c)G'(c))$$

$$\left.\cdot\left(-\frac{2}{h''(c)}\right)^{3/2} \Gamma(3/2) + o(1)\right] = 0$$

which yields

$$(3.36) \quad \psi(n) = \frac{h'''(c)}{2h''(c)\omega'(c)} - \frac{G''(c)}{G'(c)\omega'(c)} + o(1) = \frac{\omega''(c)}{(\omega(c))^2} - \frac{G''(c)}{G'(c)\omega'(c)} + o(1).$$

This establishes the expression (3.16) for $t(n)$ if $G(\lambda) \varepsilon \mathcal{G}_1$. We now turn to consideration of $G(\lambda) \varepsilon \mathcal{G}_2$.

THEOREM 3.4: *If* $G(\lambda) \varepsilon \mathcal{G}_2$, $\zeta_1 = G(l_G+) - G(l_G)$, *and* $\zeta_2 = G(u_G+) - G(u_G)$, *then*

$$(3.37) \qquad t(n) = n\frac{\int_{l_G}^{u_G} u\omega'(u)\,du}{\omega(u_G) - \omega(l_G)} + \frac{\ln\dfrac{\zeta_1(c - l_G)}{\zeta_2(u_G - c)}}{\omega(u_G) - \omega(l_G)} + o(1).$$

PROOF: As in Theorem 3.3 we must find an asymptotic expansion for the solution $t(n)$ of

$$(3.38) \quad I(n) = \int_0^\infty (\lambda - c) \exp\left\{t(n)\omega(\lambda) - n\int_\gamma^\lambda u\omega'(u)\,du\right\} dG(\lambda) = 0.$$

For $\lambda < l_G$, consider

$$(3.39) \qquad \tau_1(\lambda, n) = \frac{t(n)}{n}(\omega(\lambda) - \omega(l_G)) + \int_\lambda^{l_G} u\omega'(u)\,du.$$

Integrating by parts, and applying the mean value theorem and Theorem 3.2, we have $\limsup_{n\to\infty} \tau_1(\lambda, n) \leqq l_G(\omega(\lambda) - \omega(l_G)) + l_G\omega(l_G) - \lambda\omega(\lambda) - (l_G - \lambda)\omega(\lambda^*)$ where $\lambda < \lambda^* < l_G$. Therefore, since $\omega(\lambda)$ is increasing, we have $\limsup_{n\to\infty} \tau_1(\lambda, n) < 0$ for each $\lambda < l_G$. Hence, for each $\lambda < l_G$

$$(3.40) \qquad \frac{(\lambda - c)\exp\left\{t(n)\omega(\lambda) - n\int_\gamma^\lambda u\omega'(u)\,du\right\}}{(l_G - c)\exp\left\{t(n)\omega(l_G) - n\int_\gamma^{l_G} u\omega'(u)\,du\right\}}$$

$$= \frac{\lambda - c}{l_G - c}\exp\{n\tau_1(\lambda, n)\} \to 0,$$

as $n \to \infty$, so that

$$\int_0^{l_G+} \frac{(\lambda - c) \exp\left\{t(n)\omega(\lambda) - n \int_\gamma^\lambda u\omega'(u)\,du\right\}}{(l_G - c) \exp\left\{t(n)\omega(l_G) - n \int_\gamma^{l_G} u\omega'(u)\,du\right\}} \to \zeta_1 \tag{3.41}$$

as $n \to \infty$, by the dominated convergence theorem. This, however, is equivalent to

$$\begin{aligned} &\int_0^{l_G+} (\lambda - c) \exp\left\{t(n)\omega(\lambda) - n \int_\gamma^\lambda u\omega'(u)\,du\right\} dG(\lambda) \\ &\qquad = \zeta_1(l_G - c) \exp\left\{t(n)\omega(l_G) - n \int_\gamma^{l_G+} u\omega'(u)\,du\right\} (1 + o(1)). \end{aligned} \tag{3.42}$$

Similarly

$$\limsup_{n\to\infty} \left\{\frac{t(n)}{n} (\omega(\lambda) - \omega(u_G)) - \int_{u_G}^\lambda u\omega'(u)\,du\right\} < 0 \tag{3.43}$$

for each $\lambda > u_G$, and

$$\begin{aligned} &\int_{u_G}^\infty (\lambda - c) \exp\left\{t(n)\omega(\lambda) - n \int_\gamma^\lambda u\omega'(u)\,du\right\} dG(\lambda) \\ &\qquad = \zeta_2(u_G - c) \exp\left\{t(n)\omega(u_G) - n \int_\gamma^{u_G} u\omega'(u)\,du\right\} (1 + o(1)). \end{aligned} \tag{3.44}$$

Therefore we must determine $t(n)$ so that

$$\begin{aligned} &\zeta_1(c - l_G) \exp\left\{t(n)\omega(l_G) - n \int_\gamma^{l_G} u\omega'(u)\,du\right\} (1 + o(1)) \\ &\qquad = \zeta_2(u_G - c) \exp\left\{t(n)\omega(u_G) - n \int_\gamma^{u_G} u\omega'(u)\,du\right\} (1 + o(1)). \end{aligned} \tag{3.45}$$

Taking logarithms, we obtain (3.37) as desired.

## 4. Asymptotic characterization of the Bayes risk and the Bayes sample size.

**Theorem** 4.1: *If $F(x \mid \lambda) \in \mathfrak{F}_1$ (with $k = 1$) and $G(\lambda) \in \mathfrak{G}_1$, then the Bayes risk for fixed $n$ and $N$ is given by*

$$\begin{aligned} R(\delta^*, n, N) &= n\left((s_1 - r_1)E(\Lambda) + s_2 - r_2 + (r_1 - a_1)\int_0^c (\lambda - c)\,dG(\lambda)\right) \\ &\quad + N\left(r_1 E(\Lambda) + r_2 + (a_1 - r_1)\int_0^c (\lambda - c)\,dG(\lambda)\right) \\ &\quad + (N - n)\frac{(r_2 - a_2)(\alpha + \beta c)G'(c)}{2\alpha n} + (N - n)o\left(\frac{1}{n}\right) \end{aligned} \tag{4.1}$$

*and the Bayes sample size is*

$$n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ N^{1/2}\left(\dfrac{(r_2 - a_2)(\alpha + \beta c)G'(c)}{2\alpha A_G}\right)^{1/2} + o(N^{1/2}), & A_G > 0, \end{cases} \tag{4.2}$$

*where*

$$A_G = (s_1 - r_1)E(\Lambda) + s_2 - r_2 + (r_1 - a_1)\int_0^c (\lambda - c)\, dG(\lambda). \tag{4.3}$$

PROOF: By applying (1.20) to (1.19) we see that

$$\begin{aligned} R(\delta^*, n, N) &= n((s_1 - r_1)E(\Lambda) + s_2 - r_2) \\ &\quad + N(r_1E(\Lambda) + r_2) + (N - n)(a_1 - r_1)E\{(\Lambda - c)L(\Lambda, n)\}, \end{aligned} \tag{4.4}$$

where

$$L(\lambda, n) = E\{\delta^*(S_n) \mid \Lambda = \lambda\} = P\{S_n \leqq t(n) \mid \Lambda = \lambda\}, \tag{4.5}$$

where $t(n)$ is defined by (1.24).

Letting $\tau(n) = [t(n)]$, where $[x]$ indicates the largest integer less than equal to $x$, we may write $\tau(n) = cn + \varphi(n)$, where $\varphi(n) = O(1)$ by Theorem 3.3. Now, noting that $F^{(n)}(m \mid \lambda) \to 0$ as $\lambda \to b$ for $m < b$, we may apply Theorem 2.5 to obtain

$$\begin{aligned} L(\lambda, n) &= (\tau(n) + 1)r^{(n)}(\tau(n) + 1) \\ &\quad \cdot \int_\lambda^b \frac{t^{\tau(n)}}{(\alpha + \beta t)^{\tau(n)+1}} \exp\left\{-n\int_0^t \frac{\alpha\, du}{\alpha + \beta u}\right\} dt, \end{aligned} \tag{4.6}$$

for values of $n$ large enough so that $0 < \tau(n) < b$. Hence

$$\begin{aligned} E[(\Lambda - c)L(\Lambda, n)] &= (\tau(n) + 1)r^{(n)}(\tau(n) + 1)\int_0^{b+} (\lambda - c)\int_\lambda^b \\ &\quad \cdot \frac{t^{\tau(n)}}{(\alpha + \beta t)^{\tau(n)+1}} \exp\left\{-n\int_0^t \frac{\alpha\, du}{(\alpha + \beta u)}\, du\right\} dt\, dG(\lambda) \\ &= (\tau(n) + 1)r^{(n)}(\tau(n) + 1)\int_0^b \frac{t^{\tau(n)}}{(\alpha + \beta t)^{\tau(n)+1}} \\ &\quad \cdot \exp\left\{-n\int_0^t \frac{\alpha\, du}{\alpha + \beta u}\right\}\int_0^t (\lambda - c)\, dG(\lambda)\, dt. \end{aligned} \tag{4.7}$$

As before, let

$$h(t) = c\ln\left(\frac{t}{\alpha + \beta t}\right) - \int_0^t \frac{\alpha\, du}{\alpha + \beta u}. \tag{4.8}$$

Then

$$E[(\Lambda - c)L(\Lambda, n)] = (\tau(n) + 1)r^{(n)}(\tau(n) + 1)I(n) \tag{4.9}$$

where

$$I(n) = \int_0^b K(t, n) \exp\{nh(t)\}\, dt, \tag{4.10}$$

and

$$K(t, n) = \frac{t^{\varphi(n)}}{(\alpha + \beta t)^{\varphi(n)+1}} \int_0^t (\lambda - c)\, dG(\lambda). \tag{4.11}$$

Now

$$\frac{\partial K(c, n)}{\partial t} = \frac{c^{\varphi(n)}}{(\alpha + \beta c)^{\varphi(n)+1}} \left(\frac{\alpha\varphi(n) - \beta c}{c(\alpha + \beta c)}\right) \int_0^c (\lambda - c)\, dG(\lambda), \tag{4.12}$$

and

$$\begin{aligned} \frac{\partial^2 K(c, n)}{\partial t^2} &= \frac{c^{\varphi(n)}}{(\alpha + \beta c)^{\varphi(n)+1}} \\ &\cdot \left[ G'(c) + \frac{\alpha^2(\varphi^2(n) - \varphi(n)) - 4\alpha\beta c\varphi(n) + 2\beta^2 c^2}{c^2(\alpha + \beta c)^2} \int_0^c (\lambda - c)\, dG(\lambda)\right], \end{aligned} \tag{4.13}$$

so that

$$K(t, n) = K(c, n) + (t - c)\frac{\partial}{\partial t} K(c, n) + \frac{1}{2}\frac{\partial^2}{\partial t^2} K(c, n) + R(t, n) \tag{4.14}$$

where, by an argument similar to that used in Theorem 3.3,

$$R(t, n) = o((t - c)^2)$$

uniformly in $n$. Furthermore by Lemma 3.1

$$I(n) = \int_{c-\epsilon}^{c+\epsilon} K(t, n) \exp\{nh(t)\}\, dt + o(\exp\{nh(c)\} n^{-m}) \tag{4.15}$$

for all $m \geqq 0$. Hence, substituting (4.14) in (4.15), treating the remainder of (4.14) as was done in Theorem 3.3, and applying Lemma 3.2, we obtain

$$\begin{aligned} I(n) = \frac{\exp\{nh(c)\}}{\sqrt{n}} \left(-\frac{2}{h^{(2)}(c)}\right)^{1/2} \Bigg\{ & K(c, n)\, \Gamma\left(\frac{1}{2}\right) \\ &+ \frac{\Gamma\left(\frac{5}{2}\right)}{n} \left[ K(c, n) \left(-\frac{2}{h^{(2)}(c)}\right)^2 \left[\frac{h^{(4)}(c)}{4!} - \frac{5(h^{(3)}(c))^2}{72h^{(2)}(c)}\right] \right. \\ &\left. + \frac{\partial K(c, n)}{\partial t}\left(-\frac{2}{h^{(2)}(c)}\right)^2 \left(\frac{h^{(3)}(c)}{3!}\right) + \frac{1}{3}\frac{\partial^2 K(c, n)}{\partial t^2}\left(-\frac{2}{h^{(2)}(c)}\right)\right] + o\left(\frac{1}{n}\right)\Bigg\}. \end{aligned} \tag{4.16}$$

From (4.8) we have

$$h^{(2)}(c) = -\frac{\alpha}{c(\alpha + \beta c)},$$

$$(4.17) \qquad h^{(3)}(c) = \frac{2\alpha^2 + 4\alpha\beta c}{c^2(\alpha + \beta c)^2},$$

$$h^{(4)}(c) = -6\left(\frac{\alpha^3 + 3\alpha^2\beta c + 3\alpha\beta^2 c^2}{c^3(\alpha + \beta c)^3}\right),$$

and hence

$$I(n) = \left(\frac{2\pi c(\alpha + \beta c)}{\alpha n}\right)^{1/2} \exp\{nh(c)\} \frac{c^{\varphi(n)}}{(\alpha + \beta c)^{\varphi(n)+1}}$$

$$(4.18) \qquad \cdot \int_0^c (\lambda - c)\, dG(\lambda) \left\{1 + \frac{1}{n}\left(\frac{\alpha(\varphi^2(n) + \varphi(n))}{2c(\alpha + \beta c)} + \frac{\alpha^2 + \alpha\beta c + \beta^2 c^2}{12\alpha c(\alpha + \beta c)} + \frac{G'(c)}{2\alpha \int_0^c (\lambda - c)\, dG(\lambda)}\right) + o\left(\frac{1}{n}\right)\right\}.$$

Furthermore by (4.8)

$$(4.19) \qquad \exp\{nh(c)\} = \begin{cases} \left(\dfrac{c}{\alpha + \beta c}\right)^{cn} \left(\dfrac{\alpha}{\alpha + \beta c}\right)^{\frac{\alpha n}{\beta}}, & \beta \neq 0, \\ \left(\dfrac{c}{\alpha}\right)^{cn} \exp\{-cn\}, & \beta = 0, \end{cases}$$

and from (2.16) we have

$$(cn + \varphi(n) + 1) r^{(n)}(cn + \varphi(n) + 1)$$

$$= \begin{cases} \dfrac{\beta^{cn+\varphi(n)+1}\, \Gamma\left(\dfrac{n\alpha}{\beta} + cn + \varphi(n) + 1\right)}{(cn + \varphi(n))!\, \Gamma\left(\dfrac{n\alpha}{\beta}\right)}, & \beta > 0, \\ \dfrac{(n\alpha)^{cn+\varphi(n)+1}}{(cn + \varphi(n))!}, & \beta = 0, \\ \dfrac{(-\beta)^{cn+\varphi(n)+1} (nb)!}{(cn + \varphi(n))!\, (nb - cn - \varphi(n) - 1)!}, & -\dfrac{\alpha}{\beta} = b > 0 \end{cases}$$

$$(4.20) \qquad = \begin{cases} c^{-(cn+\varphi(n))+1/2} (\alpha + \beta c)^{\frac{n\alpha}{\beta} + cn + \varphi(n) + 1/2} \alpha^{-\frac{n\alpha}{\beta} + 1/2} \\ \cdot \sqrt{\dfrac{n}{2\pi}} \left[1 - \dfrac{\alpha(\varphi^2(n) + \varphi(n))}{2nc(\alpha + \beta c)} - \dfrac{\alpha^2 + \beta^2 c^2 + \alpha\beta c}{12n\alpha c(\alpha + \beta c)} + 0\left(\dfrac{1}{n}\right)\right], & \beta \neq 0, \\ \exp\{cn\}\, \alpha^{cn+\varphi(n)+1} \sqrt{\dfrac{n}{c\pi}}\, c^{-(cn+\varphi(n)+1/2)} \\ \cdot \left(1 - \dfrac{\varphi^2(n) + \varphi(n)}{2nc} - \dfrac{1}{12cn} + o\left(\dfrac{1}{n}\right)\right), & \beta = 0, \end{cases}$$

by Stirling's formula, using the form $\ln n! = (n + \frac{1}{2}) \ln n + \frac{1}{2} \ln 2\pi - n + 1/12n + o(1/n)$. Combining (4.19) and (4.20), we obtain

$$
\begin{aligned}
(N - n)(a_1 - r_1)E\{(\Lambda - c)L(\Lambda, n)\} &= N(a_1 - r_1) \int_0^c (\lambda - c)\, dG(\lambda) \\
&+ \frac{(r_2 - a_2)(\alpha + \beta c)}{2n\alpha} G'(c) + n(r_1 - a_1) \int_0^c (\lambda - c)\, dG(\lambda) \qquad (4.21) \\
&+ \frac{(a_2 - r_2)(\alpha + \beta c)G'(c)}{2\alpha} + (N - n)\, o\left(\frac{1}{n}\right),
\end{aligned}
$$

which upon substitution in (4.4) yields (4.1).

It is easily verified that, as long as there are sets with positive probability on each side of $c$, the Bayes decision rule leads to an incorrect decision with positive probability whenever $n$ is finite. Hence, for each finite $n$

$$E[(\Lambda - c)L(\Lambda, n)] > \int_0^c (\lambda - c)\, dG(\lambda). \qquad (4.22)$$

By (4.21), however, $E[(\Lambda - c)L(\Lambda, n)] \to \int_0^c (\lambda - c)\, dG(\lambda)$, as $n \to \infty$. Suppose that the Bayes sample size $n^*(N) = O(1)$ as $N \to \infty$. Then there exists an integer $m$ such that

$$E\{(\Lambda - c)L(\Lambda, n^*(N))\} > E\{(\Lambda - c)L(\Lambda, m))\}. \qquad (4.23)$$

Referring to (4.4) and recalling that $a_1 > r_1$ by assumption we see that this implies that

$$R(\delta^*, n^*(N), N) > R(\delta^*, m, N), \qquad (4.24)$$

for all sufficiently large values of $N$. This, however, contradicts the assertion that $n^*(N)$ is the Bayes sample size. Similarly, for any subsequence $N_k \to \infty$ the assertion that $n^*(N_k) = O(1)$ leads to a contradiction. Hence $n^*(N) \to \infty$ as $N \to \infty$.

Now for simplicity of notation we write (4.1) as

$$R(\delta^*, n, N) = A_G\, n + B_G\, N + C_G(N - n) \left(\frac{1}{n} + o\left(\frac{1}{n}\right)\right), \text{as } n \to \infty. \qquad (4.25)$$

In order to characterize the Bayes sample size we must consider two cases.

(Case i; $A_G \leqq 0$): If $A_G \leqq 0$, the risk is clearly minimized by taking $n$ as large as possible (i.e., equal to $N$) since $C_G > 0$ by assumption so that $C_G((1/n^*) + o(1/n^*)) > 0$ for large $N$ since $n^*(N) \to \infty$.

(Case ii; $A_G > 0$): Let the Bayes sample size be written as

$$n^*(N) = AN^{1/2} + \xi(N) \qquad (4.26)$$

where

$$A = \left(\frac{C_G}{A_G}\right)^{1/2} \qquad (4.27)$$

and let

$$n(N) = AN^{1/2}. \tag{4.28}$$

Now as $N \to \infty$,

$$\begin{aligned} &R(\delta^*, n^*(N), N) - R(\delta^*, n(N), N) \\ &= A_G\,\xi(N) + C_G N \left(\frac{1}{AN^{1/2} + \xi(N)} - \frac{1}{AN^{1/2}}\right)(1 + o(1)) \\ &= \frac{A_G\,\xi^2(N) + o(N^{1/2}\xi(N))}{AN^{1/2} + \xi(N)}, \end{aligned} \tag{4.29}$$

which is positive for arbitrarily large values of $N$ unless $\xi(N) = o(N^{1/2})$. If this expression is positive then the risk using $n(N)$ is less than that using $n^*(N)$ which contradicts the assertion that $n^*(N)$ is the Bayes sample size. Hence $\xi(N) = o(N^{1/2})$.

**Theorem** 4.2: *If* $F(x \mid \lambda) \in \mathfrak{F}_1$ *(with* $k = 1$*)* $G(\lambda) \in \mathfrak{G}_2$ *and* $A_G$ *is defined by* (4.3), *then the Bayes sample size is given by*

$$n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ K \ln N - \dfrac{K}{2} \ln \ln N + O(1), & A_G > 0. \end{cases} \tag{4.30}$$

(The definition of $K$ is lengthy and is contained in the proof.)

**Proof**: We rewrite (4.4) as

$$\begin{aligned} R(\delta^*, n, N) &= N(r_1 E(\Lambda) + r_2) + n((s_1 - r_1)E(\Lambda) + s_2 - r_2) \\ &+ (N - n)(a_1 - r_1)\left\{\int_0^c (\lambda - c)\, dG(\lambda) + \int_0^c (c - \lambda)(1 - L(\lambda, n))\, dG(\lambda) \right. \\ &\left. + \int_c^\infty (\lambda - c) L(\lambda, n)\, dG(\lambda)\right\}. \end{aligned}$$

Now referring to (3.37) let $\tau(n) = [l(n)] = K_1 n + \varphi(n) = O(1)$ and

$$K_1 = \frac{\displaystyle\int_{l_G}^{u_G} u\omega'(u)\, du}{\omega(u_G) - \omega(l_G)} \qquad \text{(clearly } l_G < K_1 < u_G\text{)}.$$

For $\lambda < K_1$ we may apply a well known result of asymptotic expansion theory (cf. [10]) to (2.15) to obtain

$$\begin{aligned} 1 - L(\lambda, n) &= (\tau(n) + 1)r^{(n)}(\tau(n) + 1)\left(\frac{\lambda}{\alpha + \beta\lambda}\right)^{\varphi(n)} \frac{1}{\alpha n(K_1 - \lambda)} \\ &\cdot \exp\left\{n\left(K_1 \ln \frac{\lambda}{\alpha + \beta\lambda} - \int_0^\lambda \frac{\alpha\, du}{\alpha + \beta u}\right)\right\}(1 + o(1)). \end{aligned} \tag{4.31}$$

Similarly, we obtain for $\lambda > K_1$

$$L(\lambda, n) = (\tau(n) + 1)r^{(n)}(\tau(n) + 1)\left(\frac{\lambda}{\alpha + \beta\lambda}\right)^{\varphi(n)} \tag{4.32}$$
$$\cdot \frac{1}{\alpha n(\lambda - K_1)} \exp\left\{n\left(K_1 \ln\left(\frac{\lambda}{\alpha + \beta\lambda}\right) - \int_0^\lambda \frac{\alpha\, du}{\alpha + \beta u}\right)\right\}(1 + o(1)).$$

Now, for $\lambda < K_1$, $K_1 \ln(\lambda/(\alpha + \beta\lambda)) - \int_0^\lambda \alpha\, du/(\alpha + \beta u)$ is increasing in $\lambda$, and for $\lambda > K_1$ it is decreasing. Hence we have $1 - L(\lambda, n) = o(1 - L(l_G, n))$ for $\lambda < l_G$, and $L(\lambda, n) = o(L(u_G, n))$ for $\lambda > u_G$ as $n \to \infty$. Let $\lambda^*$ be whichever of $l_G$ and $u_G$ maximizes $K_1 \ln \lambda/(\alpha + \beta\lambda) - \int_0^\lambda \alpha\, du/(\alpha + \beta u)$, and let $\zeta^*$ be the weight assigned to that point by $G(\lambda)$. At this stage we discuss in detail the case where

$$K_1 \ln \frac{l_G}{\alpha + \beta l_G} - \int_0^{l_G} \frac{\alpha\, du}{\alpha + \beta u} \neq K_1 \ln\left(\frac{u_G}{\alpha + \beta u_G}\right) - \int_0^{u_G} \frac{\alpha\, du}{\alpha + \beta u}. \tag{4.33}$$

The case where equality holds can be treated in a similar manner and leads to the same Bayes sample size, as will be noted at the conclusion of the proof. By using expression (4.2) for $(\tau(n) + 1)r^{(n)}(\tau(n) + 1)$, we have

$$R(\delta^*, n, N) = N(r_1 E(\Lambda) + r_2) + n((s_1 - r_1)E(\Lambda) + (s_2 - r_2))$$
$$+ (N - n)(a_1 - r_1)\left\{\int_0^c (\lambda - c)\, dG(\lambda)\right.$$
$$+ \left[\frac{\lambda^*(\alpha + \beta K_1)}{K_1(\alpha + \beta\lambda^*)}\right]^{\varphi(n)} \frac{(\alpha + \beta K_1)\lambda^*\zeta^*(\lambda^* - c)}{(\lambda^* - K_1)\sqrt{2\pi\alpha K_1(\alpha + \beta K_1)n}} \tag{4.34}$$
$$\cdot \begin{cases} \exp\left\{n\left(K_1 \ln \dfrac{\lambda^*}{K_1} + \dfrac{\alpha + \beta K_1}{\beta} \ln\left(\dfrac{\alpha + \beta K_1}{\alpha + \beta\lambda^*}\right)\right)\right\}, & \beta \neq 0 \\ \exp\left\{n\left(K_1 \ln \dfrac{\lambda^*}{K_1} + K_1 - \lambda^*\right)\right\}, & \beta = 0 \end{cases}$$
$$\left.\cdot(1 + o(1))\right\}.$$

Let

$$\gamma(n) = \left(\frac{\lambda^*(\alpha + \beta K_1)}{K_1(\alpha + \beta\lambda^*)}\right)^{\varphi(n)} \left(\frac{(\alpha + \beta K_1)\lambda^*\zeta^*(\lambda^* - c)}{(\lambda^* - K_1)\sqrt{2\pi\alpha K_1(\alpha + \beta K_1)}}\right), \tag{4.35}$$

$$-\frac{1}{K} = \begin{cases} K_1 \ln \dfrac{\lambda^*}{K_1} + \dfrac{\alpha + \beta K_1}{\beta} \ln \dfrac{\alpha + \beta K_1}{\alpha + \beta\lambda^*}, & \beta \neq 0, \\ K_1 \ln \dfrac{\lambda^*}{K_1} + K_1 - \lambda^*, & \beta = 0, \end{cases} \tag{4.36}$$

and note that $K > 0$. We then have, from (4.34)

$$
\begin{aligned}
R(\delta^*, n, N) &= N\left(r_1 E(\Lambda) + r_2 + (a_1 - s_1)\int_0^c (\lambda - c)\, dG(\lambda)\right) \\
&\quad + n\left((s_1 - r_1)E(\Lambda) + s_2 - r_2 + (r_1 - a_1)\int_0^c (\lambda - c)\, dG(\lambda)\right) \\
&\quad + (a_1 - s_1)(N - n)\frac{\gamma(n)}{\sqrt{n}}\exp\left\{-\frac{n}{K}\right\}(1 + o(1))
\end{aligned} \tag{4.37}
$$

as $n \to \infty$. Now, by the argument given in the proof of Theorem 4.1, $n^*(N) \to \infty$ as $N \to \infty$. As in Theorem 4.1 we have two cases, (i) $A_G \leqq 0$ and (ii) $A_G > 0$. Case (i) is treated exactly as before. Case (ii), however, requires a slightly different argument as follows: Let $n(N) = K \ln N - (K/2) \ln \ln N$ and write $n^*(N) = n(N) + \xi(N)$. Noting that

$$
\frac{N - n(N)}{\sqrt{n(N)}}\exp\left\{-\frac{1}{K}n(N)\right\} \to \frac{1}{\sqrt{K}} \quad \text{as} \quad N \to \infty, \tag{4.38}
$$

we have

$$
\begin{aligned}
&R(\delta^*, n^*(N), N) - R(\delta^*, n(N), N) \\
&= A_G\,\xi(N) + (a_1 - s_1)\left(\frac{N - n(N) - \xi(N)}{\sqrt{n(N) + \xi(N)}}\,\gamma(n(N) + \xi(N))\right. \\
&\qquad \left.\cdot\exp\left\{-\frac{1}{K}(n(N) + \xi(N))\right\}(1 + o(1)) + O(1)\right).
\end{aligned} \tag{4.39}
$$

If $\xi(N)$ or any subsequence $\to +\infty$, the exponential term is bounded and the linear term becomes infinite. If $\xi(N) \to -\infty$, then

$$
\begin{aligned}
&\frac{N - n(N) - \xi(N)}{\sqrt{n(N) + \xi(N)}}\exp\left\{-\frac{1}{K}(n(N) + \xi(N))\right\} \\
&\qquad \geqq \frac{N - n(N)}{\sqrt{n(N)}}\exp\left\{-\frac{1}{K}n(N)\right\}\exp\left\{-\frac{1}{K}\xi(N)\right\}.
\end{aligned} \tag{4.40}
$$

Recalling (4.38), and noting that $\gamma(n)$ is positive and bounded away from zero, we see that the exponential term of (4.39) becomes infinite and dominates the linear term. Therefore (4.39) is positive if $\xi(N)$ or any subsequence becomes infinite, hence $\xi(N) = O(1)$.

If equality holds in (4.33), we observe that letting

$$
\begin{aligned}
\gamma(n) &= \left(\frac{l_G(\alpha + \beta K_1)}{K_1(\alpha + \beta l_G)}\right)^{\varphi(n)}\left(\frac{(\alpha + \beta K_1)l_G\,\zeta_1(c - l_G)}{(K_1 - l_G)\sqrt{2\pi\alpha K_1(\alpha + \beta K_1)}}\right) \\
&\quad + \left(\frac{u_G(\alpha + \beta K_1)}{K_1(\alpha + \beta u_G)}\right)^{\varphi(n)}\left(\frac{(\alpha + \beta K_1)u_G\,\zeta_2(u_G - c)}{(u_G - K_1)\sqrt{2\pi\alpha K_1(\alpha + \beta K_1)}}\right)
\end{aligned} \tag{4.41}
$$

leads to (4.37), and hence to (4.30).

THEOREM 4.3: *If $F(x \mid \lambda) \in \mathfrak{F}_2$ (with $\eta = 1$), $G(\lambda) \in \mathfrak{G}_1$ and $A_G$ is defined by (4.3), then the Bayes sample size is given by*

$$n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ N^{1/2}\left(\dfrac{(r_2 - a_2)cG'(c)}{2A_G}\right)^{1/2} + o(N^{1/2}), & A_G > 0. \end{cases} \tag{4.42}$$

PROOF: The method employed here is similar to that used in the proof of Theorem 4.2. By Theorem 2.6

$$1 - L(\lambda, n) = \frac{(cn + \psi(n))^n}{\Gamma(n)} \int_0^\lambda \frac{1}{t} \exp\left\{nh(t) - \frac{\psi(n)}{t}\right\} dt \tag{4.43}$$

where $h(t) = \ln(1/t) - (c/t)$. We note that this $h(t)$ satisfies the conditions of the lemmas on asymptotic expansions and a simple calculation shows

$$\begin{aligned} R(\delta^*, n, N) = {} & n\left((s_1 - r_1)E(\Lambda) + s_2 - r_2 + (r_1 - a_1)\int_0^c (\lambda - c)dG(\lambda)\right) \\ & + N\left(r_1 E(\Lambda) + r_2 + (a_1 - r_1)\int_0^c (\lambda - c)\, dG(\lambda)\right) \\ & + (a_1 - r_1)(N - n)\left(\frac{c^2G'(c)}{2n} + o\left(\frac{1}{n}\right)\right). \end{aligned} \tag{4.44}$$

This now may be written as in (4.25) and exactly the same argument proves the theorem.

THEOREM 4.4: *If $F(x \mid \lambda) \in \mathfrak{F}_2$ (with $\eta = 1$), $G(\lambda) \in \mathfrak{G}_2$ and $A_G$ is defined by (4.3), then*

$$n^*(N) = \begin{cases} N, & A_G \leqq 0, \\ K \ln N - \dfrac{K}{2} \ln \ln N + O(1), & A_G > 0. \end{cases} \tag{4.45}$$

PROOF: We define $\lambda^*$ as being whichever of $l_G$ and $u_G$ maximizes $-(K_1/\lambda) + \ln(1/\lambda)$ and let

$$-\frac{1}{K} = \ln \frac{K_1}{\lambda^*} - \left(\frac{K_1}{\lambda^*} - 1\right). \tag{4.46}$$

The remainder of the proof parallels that of Theorem 4.2.

**Acknowledgment.** The authors are indebted to Professor James Hannan for helpful discussions concerning the results in the first part of Section 2.

## REFERENCES

[1] S. MORIGUTI, "Notes on sampling inspection plans," *Statistical Application Research, Union of Japanese Scientists and Engineers*, Vol. 3 (1955), pp. 1–23.

[2] J. BREAKWELL, "The problem of testing for the fraction of defectives," *Journal of the Operations Research Society of America*, Vol. 2 (1954), pp. 59–69.

[3] M. K. Vagholkar, Unpublished Ph.D. Dissertation, London University, 1956.

[4] S. Karlin and H. Rubin, "Theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[5] P. Halmos and L. J. Savage, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Ann. Math. Stat.* Vol. 20 (1949), pp. 225–241.

[6] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1951.

[7] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, 1934.

[8] M. Loève, *Probability Theory*, D. Van Nostrand, 1955.

[9] O. Perron, "On the asymptotic approximation of functions containing large parameters," *München Acad. Sitzungs Berichte*, 1917.

[10] A. Erdélyi, *Asymptotic Expansions*, Dover, 1956.

*Added in proof:* The following two articles survey other publications concerned with the economic aspects of acceptance sampling.

[11] H. C. Hamaker, "Economic principles in industrial sampling problems: A general introduction," *Bull. International Stat. Inst.*, Vol. 33 (1951), pp. 105–122.

[12] H. C. Hamaker, "Some basic principles of sampling inspection by attributes," *Applied Stat.*, Vol. 7 (1958), pp. 149–159.

# OPTIMUM TOLERANCE REGIONS AND POWER WHEN SAMPLING FROM SOME NON-NORMAL UNIVERSES[1]

By Irwin Guttman

*McGill University*

**1. Introduction and summary.** We assume familiarity with the concepts defined in [1] and [2], where optimum $\beta$-expectation tolerance regions and their power functions were found for $k$-variate normal distributions. The method used is to reduce this problem to that of solving an equivalent hypothesis testing problem. It is the purpose of this paper to find optimum $\beta$-expectation tolerance regions for the single and double exponential distributions, and to exhibit the corresponding power functions.

Let $X = (X_1, \cdots, X_n)$ be a random sample point in $n$ dimensions, where each $X_i$ is an independent observation, distributed by some continuous probability distribution function. It is often desirable to estimate on the basis of such a sample point a region, say $S(X_1, \cdots, X_n)$, which contains a given fraction $\beta$ of the parent distribution. We usually seek to estimate the center 100 $\beta$% of the distribution and/or one of the 100 $\beta$% tails of the parent distribution.

**2. The single exponential distribution.** The probability density function of the single exponential is given by

$$f(x)\,dx = \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x-\mu)}\,dx, \qquad x \geqq \mu \tag{2.1}$$

If we wish to construct tolerance regions $S(x_1, \cdots, X_n)$ which have the ability to pick up sets on the right hand tail of (2.1), then a reasonable choice of "the measure of desirability" $Q$ is

$$dQ_{\mu,\sigma} = \frac{1}{\alpha\sigma} e^{-\frac{1}{\alpha\sigma}(y-\mu)}\,dy, \qquad y \geqq \mu \tag{2.2}$$

where $\alpha > 1$. This clearly gives more measure to sets on the right hand tail of (2.1). The problem now separates itself into three cases.

*Case I.* $\mu$ known, $\sigma$ unknown. Without loss of generality, put $\mu = 0$. We consider the analogous hypothesis testing problem. [see p. 171 [1]]. Let $X_1, \cdots, X_n$, $Y$ be independent, each $X_i$ having the distribution (2.1), and let $Y$ have the distribution (2.2), all with $\mu = 0$. If a tolerance region is desired which tends to cover the right hand tail of (2.1), then the hypothesis testing problem has the form

$$\text{Hypothesis: } \alpha = 1;\ \text{Alternative: } \alpha = \alpha_1 > 1. \tag{2.3}$$

If $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$, then it can easily be verified that $(\bar{x}, y)$ is a sufficient statistic

Received October 7, 1958.

[1] Prepared in connection with research sponsored by the Office of Naval Research, while the author was at Princeton University.

for this problem. We now apply the invariance method expressed in terms of this sufficient statistic. Consider the group $G$ of transformations given by

$$G = \left\{ \begin{matrix} \bar{x}^1 = c\bar{x} \\ y^1 = cy \end{matrix} \middle| \, c \, \varepsilon \, (0, \infty) \right\}. \tag{2.4}$$

The function $W = y/\bar{x}$ is invariant under this group, and is in fact the maximal invariant function. It is shown in Appendix 1 that the density element of $W$ is[2]

$$g(w; \alpha) \, dw = \alpha^n n^{n+1} (n\alpha + w)^{-(n+1)} \, dw. \tag{2.5}$$

In terms of $W$, the hypothesis and alternative of (2.3) are simple, and we now apply the Neyman-Pearson fundamental Lemma. Then, the most powerful test function $\phi(w)$ is based on the probability ratio

$$\frac{\alpha_1^n n^{n+1} (n\alpha_1 + w)^{-(n+1)}}{n^{n+1}(n + w)^{-(n+1)}},$$

or, as this ratio is a monotone increasing function of $w$, $\phi(w)$ is based on $W$. Hence, the most powerful invariant test function is

$$\phi(W) = \begin{cases} 1 & \text{if} \quad W > a_\beta \\ 0 & \text{if} \quad W < a_\beta \end{cases} \tag{2.6}$$

where the $a_\beta$ are chosen to give the test size $\beta$, that is

$$\int_{a_\beta}^{\infty} g(w{:}1) \, dw = \beta. \tag{2.7}$$

Because the test does not depend on $\alpha_1$, provided it is greater than 1, and because it is based on the maximal invariant function, our most powerful invariant test function is minimax, most stringent and similar of size $\beta$. From the definition of $W$ and following [1], we have that the $\beta$-expectation tolerance region which is minimax and most stringent is given by

$$S(x_1, \cdots, x_n) = [a_\beta \bar{x}, \infty). \tag{2.8}$$

Values of $a_\beta$ for $n = 1(1)20$, 40 and 60 are given in Table I, for $\beta = .99, .95, .90$ and $.75$. The power of the procedure summarized by (2.8) is discussed in Section 4.

*Case II.* $\mu$ unknown, $\sigma$ known. Let the known value of $\sigma$ be $\sigma_0$. The sufficient statistic is $(x_{(1)}, y)$, where $x_{(1)} = \min_{i=1}^n x_i$, each $X_i$ has distribution (2.1) with $\sigma = \sigma_0$, and $Y$ has the distribution (2.2) with $\sigma = \sigma_0$. Under the group of transformations

$$G = \left\{ \begin{matrix} x_{(1)}^1 = x_{(1)} + a \\ y^1 = y + a \end{matrix} \middle| \, a \, \varepsilon \, R_1 \right\} \tag{2.9}$$

[2] Inspection of $g(w; \alpha)$ will show that it is related to Snedecor's $F$ distribution with $(2, 2n)$ degrees of freedom, where $W = \alpha F$.

TABLE I

*Tolerance Factors $a_\beta$ for single exponential distributions (2.1), $\mu$ known, $\sigma$ unknown; sample size n.*

| $n$ \ $\beta$ | .75 | .90 | .95 | .99 |
|---|---|---|---|---|
| 1 | .333333 | .111111 | .052631 | .010101 |
| 2 | .309401 | .108185 | .051957 | .010076 |
| 3 | .301927 | .107232 | .051734 | .010067 |
| 4 | .298280 | .106760 | .051624 | .010063 |
| 5 | .296119 | .106478 | .051557 | .010061 |
| 6 | .294690 | .106291 | .051513 | .010059 |
| 7 | .293675 | .106158 | .051482 | .010058 |
| 8 | .292917 | .106057 | .051458 | .010057 |
| 9 | .292329 | .105980 | .051440 | .010056 |
| 10 | .291860 | .105918 | .051425 | .010055 |
| 11 | .291476 | .105867 | .051413 | .010055 |
| 12 | .291158 | .105824 | .051403 | .010055 |
| 13 | .290889 | .105789 | .051395 | .010054 |
| 14 | .290658 | .105758 | .051387 | .010054 |
| 15 | .290458 | .105731 | .051381 | .010054 |
| 16 | .290284 | .105708 | .051376 | .010054 |
| 17 | .290131 | .105688 | .051371 | .010053 |
| 18 | .289993 | .105670 | .051366 | .010053 |
| 19 | .289871 | .105653 | .051363 | .010053 |
| 20 | .289761 | .105638 | .051359 | .010053 |
| 30 | .289066 | .105546 | .051337 | .010052 |
| 40 | .288719 | .105499 | .051326 | .010052 |
| 60 | .288373 | .105453 | .051315 | .010051 |

the statistic $W = (x_{(1)} - y)/\sigma_0$ is clearly a maximal invariant for the problem (2.3), and its distribution is given by

$$h(w;\alpha)\,dw = \begin{cases} \dfrac{n}{n\alpha+1}\,e^{-nw}\,dw & \text{if } w > 0 \\[2ex] \dfrac{n}{n\alpha+1}\,e^{w/\alpha}\,dw & \text{if } w < 0. \end{cases} \tag{2.10}$$

(This is proved in appendix 2)[3]. An analysis similar to that above shows that, for ability to pick up the right hand tail of (2.1), a minimax and most stringent tolerance region of $\beta$-expectation is

$$S(x_1, \cdots, x_n) = [x_{(1)} - b_\beta\sigma_0, \infty), \tag{2.11}$$

[3] Inspection of $h(w;\alpha)$ will show that it is a weighted combination of two densities that are simply related to $\chi^2$ with 2 degrees of freedom, where $\chi_2^2 = \alpha nW$ for $W > 0$, and $\alpha\chi_2^2 = -2W$ for $W < 0$.

TABLE II

*Tolerance Factors $b_\beta$ for single exponential distribution $\mu$ unknown, $\sigma$ known, sample size n*

| $n$ \ $\beta$ | .75 | .90 | .95 | .99 |
|---|---|---|---|---|
| 1 | .693147 | 1.60944 | 2.30258 | 3.91202 |
| 2 | .143841 | .601986 | .948560 | 1.75328 |
| 3 | .000000 | .305430 | .536479 | 1.07296 |
| 4 | −.064538 | .173287 | .346574 | .748932 |
| 5 | −.105360 | .102165 | .240794 | .562681 |
| 6 | −.133531 | .059446 | .174970 | .443209 |
| 7 | −.154151 | .031878 | .130899 | .360818 |
| 8 | −.169899 | .013170 | .099813 | .300993 |
| 9 | −.182321 | .000000 | .077016 | .255842 |
| 10 | −.192372 | −.010050 | .059784 | .220727 |
| 11 | −.200671 | −.018349 | .046439 | .192751 |
| 12 | −.207639 | −.025318 | .035899 | .170018 |
| 13 | −.213574 | −.031253 | .027437 | .151239 |
| 14 | −.218689 | −.036368 | .020549 | .135508 |
| 15 | −.223143 | −.040822 | .014876 | .122172 |
| 16 | −.227057 | −.044736 | .010157 | .110747 |
| 17 | −.230524 | −.048202 | .006198 | .100870 |
| 18 | −.233615 | −.051293 | .002850 | .092263 |
| 19 | −.236389 | −.054067 | .000000 | .084707 |
| 20 | −.238892 | −.056570 | −.002503 | .078032 |
| 30 | −.254892 | −.072571 | −.018503 | .039039 |
| 40 | −.262989 | −.080668 | −.026601 | .022290 |
| 60 | −.271152 | −.088831 | −.034764 | .008238 |

where the $b_\beta$ are chosen to give the region size $\beta$, that is the $b_\beta$ are such that

$$\int_{-\infty}^{b_\beta} h(w; 1)\, dw = \beta. \tag{2.12}$$

Values of $b_\beta$ for $n = 1(1)20$, 40 and 60 are given in Table II for $\beta = .99, .95, .90$ and $.75$. The power of the procedure as summarized by (2.11) is discussed in Section 4.

*Case III.* $\mu$ and $\sigma$ unknown. The sufficient statistic is given by $(x_{(1)}, s, y)$, where $x_{(1)} = \min_{i=1}^n x_i$, $y$ is the random variable with density (2.2), and $s$ is given by

$$s = (n-1)^{-1} \sum_{i=1}^{n} (x_i - x_{(1)}). \tag{2.13}$$

Under the group of transformations

$$(2.14)\qquad G = \left\{ \begin{array}{l} y^1 = cy + a \\ s^1 = cs \\ x^1_{(1)} = cx_{(1)} + a \end{array} \middle| \begin{array}{l} a \varepsilon R^1 \\ c \varepsilon (0, \infty) \end{array} \right\}$$

a maximal invariant is found to be

$$(2.15)\qquad W = \frac{x_{(1)} - y}{s(n^{-1} + 1)}.$$

The density element of $W$ is

$$(2.16)\qquad k(w; \alpha)\, dw = \begin{cases} \dfrac{n+1}{n\alpha + 1} \dfrac{dw}{[1 + (n+1)(n-1)^{-1}w]^n}, & w > 0 \\[2ex] \dfrac{n+1}{n\alpha + 1} \dfrac{dw}{[1 - (n+1)^{-1}(n-1)^{-1}\alpha^{-1}w]^n}, & w < 0. \end{cases}$$

(This is proved in Appendix 3)[4]. An analysis similar to that above shows that the minimax most stringent tolerance region of $\beta$-expectations, having ability to pick up the right hand tail of (2.1), is

$$(2.17)\qquad S(x_1, \cdots, x_n) = [x_{(1)} - c_\beta s, \infty],$$

where $c_\beta = (n^{-1} + 1)c^1_\beta$, and the $c^1_\beta$ are such that

$$\int_{-\infty}^{c^1_\beta} k(w; 1)\, dw = \beta.$$

The values of $c_\beta$ are given in Table III for $n = 1(1)20$, 40 and 60 for $\beta = .75$, .90, .95 and .99, while the power function for (2.17) is discussed in Section 4.

**3. The double exponential distribution.** The density of this function is given by

$$(3.1)\qquad \frac{1}{2\sigma} e^{-\frac{1}{\sigma}|x-\mu|}\, dx, \qquad -\infty < x < \infty$$

We discuss the case of $\mu$ known, say $\mu_0$. It is easily shown that if a sample of $n$ independent observations be drawn from (3.1), that the sampling distribution of the statistic

$$(3.2)\qquad T = \sum_{i=1}^{n} |X_i - \mu_0|$$

has the density

$$(3.3)\qquad \frac{1}{\sigma^n \Gamma(n)} t^{n-1} e^{-t/\sigma}\, dt$$

[4] Inspection of $k(w; \alpha)$ will show that it is a weighted combination of two densities that are simply related to an $F$ distribution with 2, $2(n-1)$ degrees of freedom, where $(n+1)W = F$ if $W > 0$, and $n\alpha F = -(n+1)W$ if $W < 0$.

TABLE III

*Tolerance Factors $c_\beta$ for single exponential distribution $\mu$ and $\sigma$ unknown, sample size $n$*

| $n$ \ $\beta$ | .75 | .90 | .95 | .99 |
|---|---|---|---|---|
| 2 | .166667 | 1.16667 | 2.83333 | 16.1666 |
| 3 | .000000 | .387426 | .824045 | 2.66666 |
| 4 | −.065238 | .194941 | .440551 | 1.28581 |
| 5 | −.106760 | .108976 | .280960 | .816410 |
| 6 | −.135330 | .061617 | .194695 | .585067 |
| 7 | −.156148 | .032478 | .141423 | .448641 |
| 8 | −.171978 | .013270 | .105729 | .359246 |
| 9 | −.184415 | .000000 | .080451 | .296463 |
| 10 | −.194424 | −.010056 | .061814 | .250149 |
| 11 | −.202698 | −.018366 | .047645 | .214709 |
| 12 | −.209611 | −.025349 | .036611 | .186807 |
| 13 | −.215486 | −.031293 | .027848 | .164334 |
| 14 | −.220539 | −.036418 | .020778 | .145895 |
| 15 | −.224931 | −.040881 | .014995 | .130528 |
| 16 | −.228784 | −.044802 | .010213 | .117554 |
| 17 | −.232192 | −.048275 | .006218 | .106474 |
| 18 | −.235227 | −.051371 | .002854 | .096920 |
| 19 | −.237948 | −.054148 | .000000 | .088609 |
| 20 | −.240400 | −.056654 | −.002503 | .081326 |
| 30 | −.256016 | −.072661 | −.018509 | .039838 |
| 40 | −.263878 | −.080751 | −.026609 | .022547 |
| 60 | −.271776 | −.088898 | −.034774 | .008273 |

Further, $T$ is sufficient for $\sigma$. If the tolerance region is constructed so that it has ability to pick up the center part of (3.1), a reasonable choice for the 'measure of desirability' is the measure $Q$, defined by

$$dQ = \frac{1}{2\alpha\sigma} e^{-\frac{1}{\alpha\sigma}|y-\mu_0|}\, dy, \tag{3.4}$$

where $-\infty < y < \infty$ and $\alpha$ is such that $0 < \alpha < 1$. The analogous hypothesis testing problem can now be put in the form

(3.5) Hypothesis: $\alpha = 1$ Alternative: $\alpha = \alpha_1$, $0 < \alpha_1 < 1$.

We use the principle of invariance. The maximal invariant under the group of transformations

$$G = \left\{ \begin{matrix} t' = ct \\ (y - \mu_0)' = c(y - \mu_0) \end{matrix} \;\middle|\; c \,\varepsilon\, (0, \infty) \right\} \tag{3.6}$$

is the statistic $W = |y - \mu_0|/t$, and its density element is given by

$$p(w; \alpha)\, dw = \frac{n\alpha^n}{(\alpha + w)^{n+1}}\, dw. \tag{3.7}$$

(This is proved in Appendix 4)[5]. In terms of $W$ the problem (3.5) is a simple hypothesis versus a simple hypothesis and clearly $(t, y)$ is sufficient. Applying the Neyman-Pearson Fundamental Lemma, the most powerful invariant test is

$$\phi(W) = \begin{cases} \text{if } W \leqq d_\beta \\ 0 \text{ otherwise} \end{cases} \tag{3.8}$$

The test does not depend on $\alpha_1$ (so long as $0 < \alpha_1 < 1$), and, because the test is based on the maximal invariant, it is minimax, most stringent, and similar of size $\beta$. The $d_\beta$ are chosen to give the test size $\beta$. Again following [1], we have the minimax most stringent tolerance regions of $\beta$-expectation with ability to put up the center 100 $\beta$% of (3.1) is

$$S(x_1, \cdots, x_n) = [\mu_0 - d_\beta t, \mu_0 + d_\beta t], \tag{3.9}$$

where the $d_\beta$ are such that

$$\int_0^{d_\beta} p(w; 1)\, dw = \beta. \tag{3.10}$$

Values of $d_\beta$ for $n = 1(1)20$, 40 and 60 for $\beta = .75, .90, .95$ and .99 are given in Table IV. The power of (3.9) is discussed in the next section.

**4. Formulation of the power functions.** Suppose sampling from (2.1), where A. *Case* 1. $\mu$ known, $\sigma$ unknown. For this case, the solution of the corresponding hypothesis testing problem is given by (2.6). The power of $\phi$, $P_\phi$, (see p. 170 of [1] and p. 774 of [2]) and hence of $S$ is determined by the distribution of $W$ under the alternative of (2.3). That is, we have

$$P_\phi = P_{\text{Alt.}}(W \geqq a_\beta) = \int_{a_\beta}^{\infty} g(w; \alpha_1)\, dw, \tag{4.1}$$

where $g(w; \alpha)$ is defined by (2.5), $a_\beta$ is given in Table I, and $\alpha_1 > 1$. The power measures the 'degree of confidence' we have that $S(X_1, \cdots, X_n)$ covers the right hand 100 $\beta$% of (2.1) when the desirability of covering this set is given by

$$Q_\sigma(S) = \int_s \frac{1}{\alpha\sigma} e^{-\frac{1}{\alpha\sigma}(x-\mu)}\, dx, \qquad 1 < \alpha.$$

For example, if it is 99.5% desirable to cover the right hand 90% of (2.1), then $\alpha_1 = 21.01938$ and the power is found by (4.1) using this value of $\alpha_1$. Values of the power for the regions $S$ (as given by (2.8)) are given in Table V when the desirability of the right hand 100 $\beta$% sets is .995.

[5] Inspection of $p(w; \alpha)$ will show that it is simply related to the $F$ distribution with $(2, 2n)$ degrees of freedom, where $nW = \alpha F$.

## TABLE IV

*Tolerance Factors $d_\beta$ for the double exponential distributions mean and variance unknown; sample size n*

| n \ β | .75 | .90 | .95 | .99 |
|---|---|---|---|---|
| 1 | 3.00000 | 9.00000 | 19.0000 | 98.9995 |
| 2 | 1.00000 | 2.16228 | 3.47214 | 8.99998 |
| 3 | .587401 | 1.15443 | 1.71442 | 3.64158 |
| 4 | .414213 | .778279 | 1.11474 | 2.16227 |
| 5 | .319508 | .584893 | .820564 | 1.51188 |
| 6 | .259921 | .467799 | .647549 | 1.15443 |
| 7 | .219014 | .389495 | .534127 | .930696 |
| 8 | .189207 | .333521 | .454215 | .778278 |
| 9 | .166529 | .291550 | .394951 | .668070 |
| 10 | .148698 | .258925 | .349283 | .584892 |
| 11 | .134312 | .232847 | .313032 | .519910 |
| 12 | .122462 | .211528 | .283560 | .467799 |
| 13 | .112531 | .193777 | .259155 | .425102 |
| 14 | .104090 | .178769 | .238599 | .389495 |
| 15 | .096825 | .165914 | .221055 | .359356 |
| 16 | .090507 | .154782 | .205908 | .333521 |
| 17 | .084964 | .145048 | .192700 | .311134 |
| 18 | .080060 | .136464 | .181080 | .291549 |
| 19 | .075691 | .128838 | .170780 | .274275 |
| 20 | .071773 | .122018 | .161586 | .258925 |
| 30 | .047294 | .079775 | .105014 | .165914 |
| 40 | .035265 | .059254 | .077770 | .122018 |
| 60 | .023374 | .039122 | .051196 | .079775 |

## TABLE V

*Power of β-expectation tolerance regions, $[a_\beta\bar{x}, \infty)$, when sampling from the single exponential distribution, sample size n*

| | Measure of Desirability = .995 | | | |
|---|---|---|---|---|
| $\alpha_1$ | 57.39245356 | 21.01937897 | 10.23299086 | 2.005037823 |
| n \ β | .75 | .90 | .95 | .99 |
| 1 | .9942255 | .9947417 | .9948830 | .9949873 |
| 3 | .9947577 | .9949156 | .9949614 | .9949958 |
| 5 | .9948565 | .9949496 | .9949769 | .9949975 |
| 7 | .9948982 | .9949642 | .9949837 | .9949982 |
| 10 | .9949289 | .9949751 | .9949885 | .9949989 |
| 15 | .9949527 | .9949839 | .9949928 | .9949994 |
| 30 | .9949772 | .9949924 | .9949968 | .9950000 |
| 60 | .9949897 | .9949968 | .9950000 | .9950000 |

### TABLE VI

*Power of β-expectation tolerance regions, $[x_{(1)} - b_\beta\sigma_0, \infty)$, when sampling from the single exponential distribution, sample size n*

| | Measure of Desirability = .995 | | | |
|---|---|---|---|---|
| $\alpha_1$ | 57.39245356 | 21.01937897 | 10.23299086 | 2.005037823 |
| $n$ \ $\beta$ | .75 | .90 | .95 | .99 |
| 1 | .9914372 | .9909171 | .9910976 | .9933444 |
| 3 | .9942255 | .9937556 | .9936906 | .9942980 |
| 5 | .9946996 | .9943447 | .9942490 | .9945578 |
| 7 | .9948414 | .9945995 | .9944926 | .9946791 |
| 10 | .9949202 | .9947892 | .9946772 | .9947744 |
| 15 | .9949637 | .9949042 | .9948218 | .9948512 |
| 30 | .9949907 | .9949755 | .9949524 | .9949305 |
| 60 | .9949977 | .9949938 | .9949880 | .9949712 |

### TABLE VII

*Power of β-expectation tolerance regions, $[x_{(1)} - c_\beta s, \infty)$ when sampling from the single exponential distribution, sample size n*

| | Measure of Desirability = .995 | | | |
|---|---|---|---|---|
| $\alpha_1$ | 57.39245356 | 21.01937897 | 10.23299086 | 2.005037823 |
| $n$ \ $\beta$ | .75 | .90 | .95 | .99 |
| 2 | .9935224 | .9930295 | .9930122 | .9940120 |
| 4 | .9945321 | .9941230 | .9940379 | .9944568 |
| 6 | .9947566 | .9944932 | .9943908 | .9946278 |
| 8 | .9948420 | .9946794 | .9945693 | .9947184 |
| 10 | .9948851 | .9947891 | .9946772 | .9947744 |
| 15 | .9949337 | .9949021 | .9948218 | .9948512 |
| 30 | .9949724 | .9949719 | .9949525 | .9949305 |
| 60 | .9949912 | .9949912 | .9949881 | .9949712 |

*Case* 2. $\mu$ unknown, $\sigma$ known. An analysis similar to the above shows that the power of (2.11) is given by

$$P_\phi = P_{\text{Alt.}}(W \leq b_\beta) = \int_{-\infty}^{b_\beta} h(w;\alpha_1)\,dw, \tag{4.2}$$

where $h(w;\alpha)$ is given by (2.10) and $b_\beta$ is given in Table II. Values of (4.2) for the regions (2.11) are given in Table VI.

TABLE VIII

*Power of β-expectation tolerance regions, $[\mu_0 - d_\beta t, \mu_0 + d_\beta t]$ when sampling from the double exponential distribution, sample size n*

| | Measure of Desirability = .995 | | | |
|---|---|---|---|---|
| $\alpha_1$ | .261648041 | .434587989 | .565411999 | .869175979 |
| $n$ \ $\beta$ | .75 | .90 | .95 | .99 |
| 1 | .9197804 | .9539367 | .9711014 | .9912967 |
| 3 | .9707346 | .9795429 | .9847458 | .9928455 |
| 5 | .9815020 | .9859235 | .9887008 | .9935183 |
| 7 | .9858373 | .9886565 | .9904921 | .9938755 |
| 10 | .9888911 | .9906625 | .9929161 | .9944304 |
| 15 | .9911096 | .9921757 | .9929161 | .9944304 |
| 30 | .9931575 | .9936285 | .9939683 | .9947047 |
| 60 | .9941067 | .9943259 | .9944876 | .9948496 |

*Case* 3. $\mu$ and $\sigma$ unknown. Proceeding as above, one finds that

$$P_\phi = P_{\text{Alt.}}(W \leqq c'_\beta) = \int_{-\infty}^{c'_\beta} k(w;\alpha_1)\,dw, \tag{4.3}$$

where $k(w;\alpha)$ is given by (2.16) and the values of $c'_\beta$ can be found from Table III using the relationship $c_\beta = (n^{-1}+1)c^{\frac{1}{2}}_\beta$. Values of (4.3) for 99.5% desirability of the right hand 100 $\beta$% sets are given in Table VII.

B. The Double Exponential Distribution. As before, the power of the regions (3.9) is given by the power of the test (3.8) under the alternative hypothesis of (3.5), that is by

$$P_\phi = P_{\text{Alt.}}(W \leqq d_\beta) = \int_0^{d_\beta} p(w;\alpha_1)\,dw \tag{4.4}$$

where $p(w;\alpha)$ is given by (3.7) and $d_\beta$ is tabulated in Table IV. Values of (4.4) are given in Table VIII.

**5. Acknowledgement.** The author wishes to express his thanks to Professor D. A. S. Fraser of the University of Toronto for valuable discussion and encouragement, and to Mr. A. Cseuz of the University of Alberta for doing the computations.

## APPENDIX

A. 1. Derivation of (2.5). To restate, the distribution of $Y$ is given by (2.2) with $\mu = 0$. Define $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, where the $X_i$ are independent observations from (2.1), with $\mu = 0$. It is well known that the density element of $\bar{X}$ is

$$\frac{1}{\sigma}\,\frac{n^n}{\Gamma(n)}\,e^{-\frac{n}{\sigma}\bar{x}}\,\bar{x}^{n-1}\,d\bar{x}.$$

Hence the joint density element of $\bar{x}$ and $y$ is

$$\frac{n^n}{\alpha\sigma^{n+1}\Gamma(n)} e^{-\frac{n}{\sigma}\bar{x}-\frac{y}{\alpha\sigma}} \bar{x}^{n-1}\, d\bar{x}\, dy.$$

We make the transformation $w = y/\bar{x}$, $z = y$. (The absolute value of the Jacobian is $z/w^2$.) The joint density element of $W$ and $Z$ is

$$g(w, z)\, dw\, dz = \frac{n^n}{\alpha\sigma^{n+1}\Gamma(n)} e^{-\frac{nz}{\sigma w}} e^{-\frac{z}{\alpha\sigma}} \frac{z^n}{w^{n+1}}\, dw\, dz.$$

On integrating out $z$ we have $g(w; \alpha)\, dw = \alpha^n n^{n+1}(n\alpha + w)^{-(n+1)}\, dw$. It is easily verified that $g(w; \alpha)$ is a probability density.

A. 2. Derivation of (2.10). Here the distribution of $Y$ is given by (2.2) with $\sigma = \sigma_0$. Define $X_{(1)} = \min_{i=1}^{n} X_i$, where $X_i$ are $n$ independent observations from (2.1) with $\sigma = \sigma_0$. It is well known that the density element of $X_{(1)}$ is given by

$$\frac{n}{\sigma_0} e^{-\frac{n}{\sigma_0}(x_{(1)}-\mu)}\, dx_{(1)}.$$

Let $s = n/\sigma_0(x_{(1)} - \mu)$ and $z = (y - \mu)/\alpha\sigma_0$. Then the density elements of $s$ and $z$ are respectively $e^{-s}\, ds$ and $e^{-z}\, dz$, and their joint density element is $e^{-s-z}\, ds\, dz$. Make the transformation

$$w = \frac{s}{n} - \alpha z \quad \text{and} \quad t = \frac{s}{n} + \alpha z.$$

Note that $w = (x_{(1)} - y)/\sigma_0$. The absolute value of the Jacobian is $n/2\alpha$. Hence

$$h(w, t)\, dw\, dt = \frac{n}{2\alpha} e^{-w\left(\frac{n}{2}-\frac{1}{2\alpha}\right)} e^{-t\left(\frac{n}{2}+\frac{1}{2\alpha}\right)}.$$

Integrating out $t$,

$$h(w; \alpha)\, dw = \begin{cases} \dfrac{n}{n\alpha + 1} e^{-nw}\, dw \text{ if } w > 0 \\[2ex] \dfrac{n}{n\alpha + 1} e^{\frac{w}{\alpha}}\, dw \text{ if } w < 0, \end{cases}$$

and it is easily verified that $h(w; \alpha)$ is a density.

A. 3. Derivation of (2.16). Using A. 2., it is easily seen that the density element of $z = (x_{(1)} - y)/(1 + n^{-1})$ is

$$\frac{n+1}{\sigma(n\alpha + 1)} e^{-\frac{n+1}{\sigma}z}\, dz \text{ if } z > 0$$

$$\frac{n+1}{\sigma(n\alpha + 1)} e^{\frac{n+1}{n\alpha\sigma}}\, dz \text{ if } z < 0,$$

where $\sigma$ is now unknown. The density element of

$$s = (n - 1)^{-1}\sum_{i=1}^{n} (x_i - x_{(1)})$$

is given by

$$\left(\frac{n-1}{\sigma}\right)^{n-1} \frac{1}{\Gamma(n-1)} s^{n-2} e^{-\frac{(n-1)s}{\sigma}} ds$$

([3], p. 54). Hence the joint density element of $z$ and $s$ is

$$\frac{n+1}{\sigma(n\alpha+1)} \left(\frac{n-1}{\sigma}\right)^{n-1} \frac{s^{n-2}}{\Gamma(n-1)} e^{-\frac{(n-1)s}{\sigma}-\left(\frac{n+1}{\sigma}\right)z} ds\, dz \text{ if } z > 0$$

and

$$\frac{n+1}{\sigma(n\alpha+1)} \left(\frac{n-1}{\sigma}\right)^{n-1} \frac{s^{n-2}}{\Gamma(n-1)} e^{-\frac{(n-1)s}{\sigma}} e^{\frac{n+1}{n\sigma\alpha}z} ds\, dz \text{ if } z < 0.$$

Making the transformation $w = z/s$ and $r = z$ (the absolute value of the Jacobian is $r/w^2$), the joint distribution of $w$ and $r$ becomes

$$k(w, r)\, dw\, dr = \begin{cases} \dfrac{n+1}{\sigma(n\alpha+1)} \left(\dfrac{n-1}{\sigma}\right)^{n-1} \dfrac{r^{n-1}}{w^n\Gamma(n-1)} e^{-\left(\frac{n-1}{\sigma}\right)\frac{r}{w}} e^{-\left(\frac{n+1}{\sigma}\right)r}\, dw\, dr & \text{if } w > 0 \\ \dfrac{n+1}{\sigma(n\alpha+1)} \left(\dfrac{n-1}{\sigma}\right)^{n-1} \dfrac{r^{n-1}}{w^n\Gamma(n-1)} e^{-\frac{(n-1)r}{\sigma w}} e^{\frac{(n+1)r}{\alpha n\sigma}}\, dw\, dr & \text{if } w < 0. \end{cases}$$

Integrating out $r$

$$k(w; \alpha)\, dw = \begin{cases} \dfrac{n+1}{n\alpha+1} \dfrac{dw}{[1+(n+1)(n-1)^{-1}w]^n} \text{ if } w > 0 \\ \dfrac{n+1}{n\alpha+1} \dfrac{dw}{[1-(n+1)n^{-1}(n-1)^{-1}\alpha^{-1}w]^n} \text{ if } w < 0, \end{cases}$$

and it is readily seen that $k(w; \alpha)$ is a density.

A. 4. Derivation of (3.7). Let $Y$ have the distribution (3.4) and define $T = \sum_{i=1}^n |X_i - \mu_0|$, $V = |Y - \mu_0|$, where each $X_i$ is distributed by (3.1), and so $T$ has the density (3.3). It is easily shown that $V$ has the density element

$$\frac{1}{\alpha\sigma} e^{-\frac{v}{\alpha\sigma}} dv, \qquad v \geqq 0.$$

The joint density element of $V$ and $T$ is then

$$\frac{1}{\alpha\sigma^{n+1}} \frac{t^{n-1}}{\Gamma(n)} e^{-t/\sigma} e^{-v/\alpha\sigma} dt\, dv.$$

If we let $w = v/t$ and $z = t$ (the absolute value of the Jacobian is $z$), the joint density element is

$$p(w, z)\, dw\, dz = \frac{1}{\alpha\sigma^{n+1}} \frac{z^n}{\Gamma(n)} e^{-z/\sigma} e^{-\frac{zw}{\alpha}} dw\, dz.$$

Integrating over $z$

$$p(w; \alpha)\, dw = \frac{n\alpha^n}{(\alpha + w)^{n+1}}\, dw, \qquad w > 0,$$

and it is easily verified that $p(w; \alpha)$ integrates to 1.

## REFERENCES

[1] D. A. S. Fraser and Irwin Guttman, "Tolerance regions," *Ann. Math. Stat.*, Vol. 27 (1956), p. 162.

[2] Irwin Guttman, "On the power of optimum tolerance regions when sampling from normal distributions," *Ann. Math. Stat.*, Vol. 28 (1957), p. 773.

[3] P. V. Sukhatme, "Tests of signifince for samples of the $\chi^2$-population with two degrees of freedom," *Ann. Eugenics*, Vol. 8 (1937), p. 52.

# PROPERTIES OF MODEL II—TYPE ANALYSIS OF VARIANCE TESTS, A: OPTIMUM NATURE OF THE $F$-TEST FOR MODEL II IN THE BALANCED CASE[1, 2]

By Leon H. Herbach

*College of Engineering, New York University*[3]

**1. Summary.** A distribution analogous to the canonical distribution used in testing the general linear hypothesis is developed for Model II analysis of variance for balanced classifications. As in the case of Model I analysis of variance, this standard distribution exhibits the sums of squares going into the analysis of variance table. By use of the standard form it is also shown that (i) all exact $F$-tests used in testing hypotheses based on balanced multiple classifications determine uniformly most powerful (u.m.p.) similar regions although they are not likelihood ratio (L.R.) tests, but (ii) in the balanced one-way classification, for all practical purposes, the test is an L.R. test, and is u.m.p. invariant. An *exact* $F$-test exists when we have a sum of squares, $S_1$ distributed as $(k + \sigma_0^2)$ times a chi-square variate, where $k > 0$, independently of $S_2$, which is distributed as $k$ times a chi-square variate. The test is then to reject the hypothesis that $\sigma_0^2 = 0$ whenever $S_1/S_2$ is greater than some suitably chosen number, $c$. As a corollary to property (i) it is shown that "of all invariant tests of $\sigma_0^2 = 0$ against $\sigma_0^2 > 0$ whose power is a function of $\sigma_0^2/(k + \sigma_0^2)$ only, the test $S_1/S_2 > c$ is most powerful, providing $S_1$ and $S_2$, as defined above can be found."

**2. Notation and terminology.** We use the notation $p_\theta(x)$ for the probability density function (p.d.f.) of the vector-valued random variable, $X$, which depends on the vector-valued parameter $\theta \varepsilon \Omega$, where $\Omega$ will always represent the unrestricted parameter space. This notation is generic so that $p$ may not be the same density each time it appears. The difference in functional form is indicated by the change in variable. The actual form will always be clear from the context. This same generic notation will be used for constants; $c$ will usually be a constant, not necessarily the same one each time it appears. It will be clear from the context when $c$ is not a constant. The subspace of $\Omega$ specified by the hypothesis being tested will be denoted by $\omega$. No confusion will be caused when dealing with the hypothesis $H: \theta \varepsilon \omega$ if we sometimes speak of $\omega$ rather than $H$ as the hypothesis. By a test of an hypothesis we mean any measurable function $\varphi(x)$ with the property that $0 \leqq \varphi(x) \leqq 1$. When $X$ is observed to take on the value $x$ one rejects $H$ with probability $\varphi(x)$.

**3. Introduction.** In Model II (components of variance model) analysis of

Received May 8, 1958; revised January 15, 1959.

[1] Part of the author's doctoral dissertation submitted to Columbia University [3].

[2] Work partially sponsored by the Office of Naval Research.

[3] Scientific Paper No. 6a, Engineering Statistics Laboratory.

variance, the following stochastic model is assumed in the case of a two-way classification with $K$ observations per cell:

$$X_{ijk} = \mu + e_i^A + e_j^B + e_{ij}^{AB} + e_{ijk}, \quad (3.1)$$
$$i = 1, \cdots, I; \quad j = 1, \cdots, J; \quad k = 1, \cdots K,$$

where $X_{ijk}$ is the $k$th measurement on the $(i, j)$th cell, $\mu$ the main effect is assumed a constant and the "components" $e_i^A$, $e_j^B$, $e_{ij}^{AB}$, $e_{ijk}$ are normally and independently distributed (NID) with means zero and variances $\sigma_a^2$, $\sigma_b^2$, $\sigma_{ab}^2$, $\sigma_e^2$ respectively. These will be referred to as the Model II assumptions. If, as here, one has the same number of observations in each cell, the classification is called *balanced*, otherwise unbalanced.

The corresponding model for Model I (general linear hypothesis model) analysis of variance is given by (3.1) where it is now assumed that in addition to $\mu$, $e_i^A$, $e_j^B$ and $e_{ij}^{AB}$ are also constants, and $e_{ijk}$ are the only random variables, and these are NID$(0, \sigma^2)$. Furthermore it is usually assumed that $\sum_i e_i^A = \sum_j e_j^B = \sum_i e_{ij}^{AB} = \sum_j e_{ij}^{AB} = 0$. These equations for the effects may be assumed without loss of generality in Model I but would violate the assumed independence in Model II. The usual theoretical procedure in setting up any Model I hypothesis, say $H_0 : a_i = 0$ $(i = 1, \cdots, I)$, is to find the likelihood ratio test of the hypothesis. This gives the usual $F$-test. In addition to having the backing of the intuitive appeal of the likelihood ratio test, the resulting $F$-test has been shown by Hsu [4], [5], Wald [14], Wolfowitz [16] and others to have many optimum properties.

Analysis of variance, to many, also means a technique of calculating the analysis of variance table given in Table 3.1, where

$$\begin{aligned}
S_1 &= IJK\,(X_{\cdots} - \mu)^2 \\
S_2 &= JK\sum_i (X_{i\cdot\cdot} - X_{\cdots})^2 \\
S_3 &= IK\sum_j (X_{\cdot j\cdot} - X_{\cdots})^2 \qquad (3.2)\\
S_4 &= K\sum_i\sum_j (X_{ij\cdot} - X_{i\cdot\cdot} - X_{\cdot j\cdot} + X_{\cdots})^2 \\
S_5 &= \sum_i\sum_j\sum_k (X_{ijk} - X_{ij\cdot})^2.
\end{aligned}$$

TABLE 3.1

*Analysis of Variance Table for a Balanced Two-way Classification*

| Source | d.f. | S.S. | m.s. | $E$ (mean square) |
|---|---|---|---|---|
| mean | $\nu_1 = 1$ | $S_1$ | $S_1$ | $\lambda_1$ |
| $A$ effect | $\nu_2 = (I - 1)$ | $S_2$ | $S_2/\nu_2$ | $\lambda_2$ |
| $B$ effect | $\nu_3 = (J - 1)$ | $S_3$ | $S_3/\nu_3$ | $\lambda_3$ |
| $AB$ interaction | $\nu_4 = (I - 1)(J - 1)$ | $S_4$ | $S_4/\nu_4$ | $\lambda_4$ |
| error | $\nu_5 = IJ(K - 1)$ | $S_5$ | $S_5/\nu_5$ | $\lambda_5$ |

The mean row and $E$ (mean square) column do not always appear in the usual analysis of variance table and will be explained later. The statistic used in Model I to test $H_0$ is $(\nu_5 S_2)/(\nu_2 S_5)$, which is distributed as $F$ with $\nu_2$ and $\nu_5$ degrees of freedom.

The procedure in Model II is to set up the analysis of variance table that is *used in Model I*, and then to add a column which gives the expected mean squares. One then notes that the five mean squares are always independently distributed and that $S_i/(\nu_i\lambda_i)$ is distributed as $\chi^2$ with $\nu_i$ degrees of freedom. Using the fact that the expected mean squares are

$$\begin{aligned} \lambda_1 &= \sigma_e^2 + K\sigma_{ab}^2 + JK\sigma_a^2 + IK\sigma_b^2 \\ \lambda_2 &= \sigma_e^2 + K\sigma_{ab}^2 + JK\sigma_a^2 \\ \lambda_3 &= \sigma_e^2 + K\sigma_{ab}^2 + IK\sigma_b^2 \\ \lambda_4 &= \sigma_e^2 + K\sigma_{ab}^2 \\ \lambda_5 &= \sigma_e^2 \end{aligned} \tag{3.3}$$

we have, under the hypothesis of no Model II $A$ effect $(H_0': \sigma_a^2 = 0)$, that $\lambda_2 = \lambda_4$ and $(\nu_4 S_2)/(\nu_2 S_4)$ is distributed as $F$ with $\nu_2$ and $\nu_4$ degrees of freedom. This, in fact, is the $F$-statistic used in testing $H_0'$. All exact $F$-tests used in Model II are obtained by taking ratios of mean squares which have equal $\lambda$'s under the hypothesis, *whenever there are equal* $\lambda$'s. No attempt has been made previously to show that these tests are optimum or to even show they are likelihood ratio (L.R.) tests, which they sometimes are not. Two of the purposes of this paper are to derive optimum properties for some tests of Model II hypotheses and to show that in this model the analysis of variance table can be obtained without borrowing it from Model I.

**4. Some useful lemmas for a certain matrix.** The following $n \times n$ matrix plays an important role in what follows:

$$A = \begin{Bmatrix} a+b & a & a \cdots & a \\ a & a+b & a \cdots & a \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a & a & a \cdots & a+b \end{Bmatrix}, \tag{4.1}$$

where $a$ and $b$ are either *scalars or square matrices* of the same size. Since $A$ is a function of $a + b$ and $a$ only, the notation

$$A = (a + b\backslash a)^4 \tag{4.2}$$

shall be used.

[4] It may be noted that for $a$, $b$ scalars, $A = b\mathcal{I} + a\mathcal{I}^*$, where $\mathcal{I}$ is the unit matrix and $\mathcal{I}^*$ is the matrix with all elements unity.

We shall make use of the following two lemmas:

LEMMA 4.1: *If $A$ is of the form* (4.1) *then the determinant of $A$ satisfies*

$$| A | = | b + na || b |^{n-1}$$

*where $| D |$ means the determinant of the matrix $D$.*

The proof is exactly the same as that given by Wilks ([15], p. 109) for the case in which $a$ and $b$ are scalars.

LEMMA 4.2: *If $a$, $b$ are real numbers with $b(b + na) \neq 0$,*

$$A^{-1} = \frac{1}{(b + na)b} ([b + (n - 1)a] \backslash -a). \tag{4.3}$$

**5. Standard form for the balanced two-way classification.** Consider the two-way Model II classification with $K$ observations per cell, given by (3.1). Let the transpose of the observation vector $X$ be

$$\begin{aligned} X' = (X_{111}, X_{211}, X_{311}, \cdots, X_{I11};\ X_{121}, X_{221}, \cdots, X_{I21}; \cdots ;\\ X_{1J1}, X_{2J1}, X_{3J1}, \cdots, X_{IJ1};\ X_{112}, X_{212}, X_{312}, \cdots; X_{I12};\\ X_{122}, X_{222}, \cdots, X_{I22}; \cdots ;\ X_{IJK}, X_{2JK}, X_{3JK}, \cdots, X_{IJK}), \end{aligned} \tag{5.1}$$

that is, the triplets $i, j, k$ are ordered so that

$$\begin{aligned} (i, j, k) &\text{ precedes } (i', j, k) &&\text{if } i < i',\\ (i, j, k) &\text{ precedes } (i', j', k) &&\text{if } j < j',\\ (i, j, k) &\text{ precedes } (i', j', k') &&\text{if } k < k'. \end{aligned}$$

Let $\mathfrak{T}$ be the covariance matrix of $X$ and $N = IJK$. Then it is known that there exists an orthogonal matrix $D$ with the following properties: (a) its first row is $N^{-\frac{1}{2}}\delta'$, where

$$\delta' = (1, 1, \cdots, 1), \tag{5.2}$$

a $1 \times N$ row vector, (b) the covariance matrix of $Z = DX$ is $D \mathfrak{T} D' = \Lambda = \text{diag.} (\lambda_1, \cdots, \lambda_N)$, $\lambda_i > 0$ and (c) the $\lambda$'s are the roots of the characteristic equation $| \mathfrak{T} - \lambda \mathcal{I} | = 0$, where $\mathcal{I}$ is the identity matrix.

We shall now find these $\lambda$'s in terms of $\sigma_a^2$, $\sigma_b^2$, $\sigma_{ab}^2$ and $\sigma_e^2$. Now

$$| \underset{K\times K}{\mathfrak{T}} - \lambda \mathcal{I} | = | \mathcal{A} \backslash \mathcal{B} |$$

where

$$\begin{aligned} \mathcal{A} &= (A_1 \backslash B_1)\\ \underset{J\times J}{\mathcal{B}} &= (A_2 \backslash B_2)\\ A_1 &= (\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \sigma_e^2 - \lambda \backslash \sigma_b^2)\\ B_1 &= (\sigma_a^2 \backslash 0) = B_2\\ \underset{I\times I}{A_2} &= (\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 \backslash \sigma_b^2). \end{aligned} \tag{5.3}$$

It should be noted that $\mathfrak{T}$ is an $N \times N$ matrix of scalars, but is a $K \times K$ matrix when the elements are submatrices ($\mathfrak{A}$'s and $\mathfrak{B}$'s). Repeated use of Lemma 4.1 yields

$$\begin{aligned}
|\mathfrak{T} - \lambda \mathfrak{g}| &= |(\mathfrak{A} - \mathfrak{B}) + K\mathfrak{B}\,\|\,\mathfrak{A} - \mathfrak{B}\,|^{K-1} \\
&= |\,A_1 + (K-1)A_2\backslash KB_2\,|\cdot|\,A_1 - A_2\backslash 0|^{K-1} \\
&= |\,A_1 + (K-1)A_2 + K(J-1)B_2\,|\cdot|\,A_1 + (K-1)A_2 \\
&\quad - KB_2\,|^{J-1}\,|\,A_1 - A_2\,|^{J(K-1)} \\
&= D_1 \cdot D_2 \cdot D_3\,, \text{ say,}
\end{aligned}$$

where

$$\begin{aligned}
D_1 &= |\,K(\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2) + \sigma_e^2 + K(J-1)\sigma_a^2 - \lambda\backslash K\sigma_b^2\,| \\
&= |\,K\sigma_{ab}^2 + \sigma_e^2 + JK\sigma_a^2 + IK\sigma_b^2 - \lambda\,\|\,K\sigma_{ab}^2 + \sigma_e^2 + JK\sigma_a^2 - \lambda\,|^{I-1} \\
D_2 &= |\,K(\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2) + \sigma_e^2 - \lambda - K\sigma_a^2\backslash K\sigma_b^2\,|^{J-1} \\
&= |\,K\sigma_{ab}^2 + IK\sigma_b^2 + \sigma_e^2 - \lambda\,|^{J-1}\cdot|\,K\sigma_{ab}^2 + \sigma_e^2 - \lambda\,|^{(I-1)(J-1)} \\
D_3 &= |\,\sigma_e^2 - \lambda\backslash 0\,|^{J(K-1)} = |\,\sigma_e^2 - \lambda\,|^{IJ(K-1)}.
\end{aligned}$$

Therefore the values of the $N = IJK$ characteristic roots are

$$\begin{aligned}
&\sigma_e^2 + K\sigma_{ab}^2 + IK\sigma_b^2 + JK\sigma_a^2 &&= \lambda_1\,, \text{ say with multiplicity } 1 \\
&\sigma_e^2 + K\sigma_{ab}^2 + JK\sigma_a^2 &&= \lambda_2\,, \text{ say with multiplicity } (I-1) \\
&\sigma_e^2 + K\sigma_{ab}^2 + IK\sigma_b^2 &&= \lambda_3\,, \text{ say with multiplicity } (J-1) \\
&\sigma_e^2 + K\sigma_{ab}^2 &&= \lambda_4\,, \text{ say with multiplicity } (I-1)(J-1) \\
&\sigma_e^2 &&= \lambda_5\,, \text{ say with multiplicity } IJ(K-1).
\end{aligned}$$

These $\lambda$'s are the same as the ones defined by (3.3). The orthogonality of $D$, the property that the first row of $D$ is $N^{-\frac{1}{2}}\delta'$ and the fact that $EX_{ijk} = \mu$ imply that

$$\text{(5.4)} \qquad \begin{aligned} EZ_{111} &= \sqrt{N}\mu \\ EZ_{ijk} &= 0, \qquad \text{for } (i, j, k) \neq (1, 1, 1). \end{aligned}$$

After the dissertation [3] was defended but before this paper was prepared, Dr. Howard Levene called the author's attention to the work of Nelder [11] whose method for obtaining the latent roots of a special case of matrices of the form (4.1) can be generalized to find our eigenvalues. However, it is felt that the algorithm, using Lemma 4.1 is more convenient, especially when higher multiple classifications are treated.

Let $\zeta = EZ$, the vector given in (5.4). We have shown that the vector variable $X$ defined by (3.1) which is distributed as $N(\mu\delta, \mathfrak{T})$ where $\mathfrak{T} = (\mathfrak{A}\backslash\mathfrak{B})$

and $\mathfrak{A}$, $\mathfrak{B}$ are defined by (5.3) when $\lambda = 0$, may be transformed by an orthogonal matrix to yield a variable $Z$ which has the following density:

$$(5.5)\quad \frac{|\Lambda|^{-\frac{1}{2}}}{(2\pi)^{N/2}} e^{-\frac{1}{2}[(z-\zeta)'\Lambda^{-1}(z-\zeta)]} = \frac{|\Lambda|^{-\frac{1}{2}}}{(2\pi)^{IJK/2}} \exp\left\{-\frac{1}{2}\left(\frac{s_1}{\lambda_1} + \frac{s_2}{\lambda_2} + \frac{s_3}{\lambda_3} + \frac{s_4}{\lambda_4} + \frac{s_5}{\lambda_5}\right)\right\},$$

where $\lambda_1, \cdots, \lambda_5$ are given by (3.3) and

$$(5.6)\quad \begin{aligned} s_1 &= (z_{111} - \sqrt{N}\mu)^2 \\ s_2 &= \sum_{i=2}^{I} z_{i11}^2 \\ s_3 &= \sum_{j=2}^{J} z_{1j1}^2 \\ s_4 &= \sum_{i=2}^{I}\sum_{j=2}^{J} z_{ij1}^2 \\ s_5 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=2}^{K} z_{ijk}^2 \end{aligned}$$

The reader should note that $s_1$ is not a statistic, since it contains $\mu$. This particular expression for $s_1$ is used because of the symmetry it gives to (5.5), which we shall refer to as the Model II standard form of the probability density for the case of a balanced two way classification with $K$ observations per cell. Note that (3.2) implies that

$$(5.7)\quad \lambda_1 = \lambda_2 + \lambda_3 - \lambda_4,$$

a fact which will be used later.

For completeness the Tang canonical form [13] of the joint density of (3.1) when the Model I assumptions are made will now be given. Then

$$X_{ijk}:\mathrm{NID}(\mu_{ij}, \sigma^2)$$

where

$$\mu_{ij} = \mu + e_i^A + e_j^B + e_{ij}^{AB}.$$

Tang showed that there exists an orthogonal $N \times N$ matrix $D$ whose first row is $N^{-\frac{1}{2}}\delta'$ for which $Z = DX$ has the density,

$$(5.8)\quad \begin{aligned} \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\Bigg[-\frac{1}{2\sigma^2}\Bigg\{&(z_{111} - \sqrt{N}\mu)^2 \\ &+ \sum_{i=2}^{I}(z_{i11} - a_i^A)^2 + \sum_{j=2}^{J}(z_{1j1} - a_j^B)^2 \\ &+ \sum_{i=2}^{I}\sum_{j=2}^{J}(z_{ij1} - a_{ij}^{AB}) + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=2}^{K} z_{ijk}^2\Bigg\}\Bigg] \end{aligned}$$

where the $a_i^A(a_j^B, a_{ij}^{AB})$ are linear combinations of the $e_i^A(e_j^B, e_{ij}^{AB})$ such that $a_i^A(a_j^B, a_{ij}^{AB})$ are zero if and only if all $e_i^A(e_j^B, e_{ij}^{AB})$ are zero. It should be noted

that in Model I one transforms to change the means while in Model II one does so to change the covariance matrix. It can be shown [3] that the same orthogonal transformation, $D$, could be used in both models, so that we are justified in using the same letter $Z$ in (5.6) and (5.8). Also, Tang showed for Model I, that $S_1, \cdots, S_5$ as given in (3.2) and (5.6) are the same, except that (3.2) expresses the $S$'s in terms of original variates while (5.6) does so in terms of transformed variates. Since the transformations are the same in both models this shows that the sums of squares in (3.2) and (5.6) are the same in Model II. From this one can argue that the standard distribution (5.5) can be obtained from the analysis of variance table above since it is known that all rows are independently distributed. This was not done because we do not yet know what the properties of the tests based on Table 3.1 are. We propose to get the table, tests and optimum properties of the tests from (5.5), the standard form, which is easier to handle than the density of $X$, although all tests of hypotheses based on $X$ can be transformed to tests based on $Z$. It should be noted that (5.5) is the density of $Z$ although it is written in terms of the $s$'s. From (5.5) and (5.6) it is clear that $\bar{X} = N^{-\frac{1}{2}}Z_{111}$, $S_2$, $S_3$, $S_4$ and $S_5$ are independently distributed as a normal variate and four multiples of $\chi^2$ with $(I-1)$, $(J-1)$, $(I-1)(J-1)$ and $IJ(K-1)$ degrees of freedom respectively. In the sequel we shall use this latter joint density, namely,

$$(5.9)\quad p(\bar{x}, s_2, s_3, s_4, s_5) = \left(\frac{N}{2\pi\lambda_1}\right)^{\frac{1}{2}} \exp\left[-\frac{N(\bar{x}-\mu)^2}{2\lambda_1}\right] \prod_{i=2}^{5} \frac{s_i^{\frac{\nu_i-2}{2}} \exp\left(-\frac{s_i}{2\lambda_i}\right)}{(2\lambda_i)^{\nu_i/2}\Gamma(\nu_i/2)}$$

Densities (5.5) and (5.8) or (5.9) and (5.8) show clearly that under the hypothesis of no $A$ effect, $H_0'$ and $H_0$ respectively, $S_2/S_4$ and $S_2/S_5$ respectively are distributed as a multiple of $F$ with the degrees of freedom indicated by the number of standard variates in each $S$. These are the statistics indicated at the end of Section 3. All $F$-tests used to test the non-existence of certain effects can be obtained this way.

**6. Uniformly most powerful similar test for testing non-existence of main effects in the balanced one or two-way classification.** This section will be devoted to showing that the $F$-test is the u.m.p. similar test for testing $\omega:\sigma_a^2 = 0$ against $\Omega - \omega:\sigma_a^2 > 0$ when one has a balanced one or two way Model II classification. Although the hypothesis to be tested is actually $\sigma_a^2 = 0$, $\sigma_b^2 \geqq 0$, $\sigma_{ab}^2 \geqq 0$ and $\sigma_e^2 > 0$ we defer to the usual practice of not explicitly stating the other inequalities when no confusion will result. A similar statement can be made in regard to the alternative hypothesis. In the two-way classification

$$\Omega = \{\theta \mid -\infty < \mu < \infty;$$

$$(6.1)\qquad 0 < \lambda_5 \leqq \lambda_4 \leqq \lambda_2 \leqq \lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 < \infty;$$

$$\lambda_4 \leqq \lambda_3 \leqq \lambda_1\}$$

$$(6.2)\quad \omega = \{\theta \mid -\infty < \mu < \infty; \quad 0 < \lambda_5 \leqq \lambda_4 = \lambda_2 \leqq \lambda_3 = \lambda_1 < \infty\}$$

where $\theta = (\mu, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$. Since $z_{111} = \sqrt{N}\bar{x}$ it can be seen from (5.5) that a sufficient statistic under $\omega$, for the two-way classification is

$$T = (\bar{X}, S_3, S_5, U) \tag{6.3}$$

where

$$U = S_2 + S_4 \tag{6.4}$$

We first prove the following

THEOREM 6.1: *For the standard distribution of the two-way Model* II *classification* (5.5) *the statistic T defined by* (6.3) *is complete on* $\omega$, *where* $\omega$ *is determined by the hypothesis* $\sigma_a^2 = 0$ *and is defined by* (6.2).

PROOF: By the definition of completeness [7] we need to show that

$$E_\theta f(T) \underset{\theta \varepsilon \omega}{=} 0$$

implies $f(t) = 0$, (a.e.). For $\theta \varepsilon \omega$, we have, using (5.9),

$$E_\theta f(T) = c(\theta) \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} f(t) g(t, \lambda_3) h(t, \theta) \, ds_5 \, du \, ds_3 \, d\bar{x} \tag{6.5}$$

where

$$g(t, \lambda_3) = \exp\left\{ \frac{-N\bar{x}^2}{2\lambda_3} \right\} s_3^{\frac{\nu_3 - 2}{2}} u^{\frac{\nu_2 + \nu_4 - 2}{2}} s_5^{\frac{\nu_5 - 2}{2}} \tag{6.6}$$

and

$$h(t, \theta) = \exp\left\{ \frac{N\mu\bar{x}}{\lambda_3} - \frac{s_3}{2\lambda_3} - \frac{u}{2\lambda_4} - \frac{s_5}{2\lambda_5} \right\}. \tag{6.7}$$

Let $S_3^* = S_3 + N\bar{X}^2$ and $T^* = (\bar{X}, S_3^*, S_5, U)$. Changing the variable of integration in (6.5) to $t^*$ one gets for $\theta \varepsilon \omega$,

$$\begin{aligned} E_\theta f(T) = c(\theta) \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} & f^*(t^*) g^*(t^*) \\ & \exp\left\{ \frac{N\mu\bar{x}}{\lambda_3} - \frac{s_3^*}{2\lambda_3} - \frac{u}{2\lambda_4} - \frac{s_5}{2\lambda_5} \right\} ds_5 \, du \, ds_3^* \, d\bar{x} \end{aligned} \tag{6.8}$$

where

$$f^*(t^*) = \begin{cases} f(t) & \text{if } s_3^* > N\bar{x}^2 \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

and

$$g^*(t^*) = (s_3^* - N\bar{x}^2)^{\frac{\nu_3 - 2}{2}} u^{\frac{\nu_2 + \nu_4 - 2}{2}} s_5^{\frac{\nu_5 - 2}{2}}. \tag{6.10}$$

By the unicity property of the quadruple Laplace transform, (6.8) is identically zero for $\theta$ in a non-degenerate interval only if

$$f(t^*) g^*(t^*) \equiv 0 \quad \text{(a.e.)}. \tag{6.11}$$

Now $g^*(t^*) \neq 0$ (a.e.). Thus (6.11) and (6.9) imply that $f(t) \equiv 0$ (a.e.) and the theorem is proved.

Let $V$ be defined by the following 1:1 transformation

$$(6.12)\qquad \begin{array}{ll} U = S_2 + S_4 & \qquad S_2 = UV \\ & \text{or} \\ V = S_2/(S_2 + S_4) & \qquad S_4 = U(1 - V). \end{array}$$

Since, as can be seen from (5.5), $(\bar{X}, S_2, S_3, S_4, S_5)$ is sufficient under $\Omega$ then $W = (T, V)$ is also. Using (6.12) and the density of $(S_2, S_4)$ given by (5.9) we have for the density of $(U, V)$,

$$(6.13)\qquad p_\theta(u, v) = c(\theta) u^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)} e^{-\frac{u}{2\lambda_4}}, \qquad \theta \varepsilon \Omega.$$

But under $\omega$, $\lambda_2 = \lambda_4$ and (6.13) becomes for $\theta_0 \varepsilon \omega$

$$(6.14)\qquad p_{\theta_0}(u, v) = c(\theta_0) u^{\frac{\nu_2+\nu_4-2}{2}} e^{-\frac{u}{2\lambda_4}} v^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}},$$

which shows that $U$ and $V$ are independent under $\omega$. Since (5.9) and (6.4) show clearly that $(\bar{X}, S_3, S_5)$ and $(U, V)$ are always independent this means that $T$ and $V$ are independent under $\omega$ and we have

$$(6.15)\qquad p_\theta(v \mid t) = p_\theta(v) = cv^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}}, \qquad \text{for} \quad \theta \varepsilon \omega.$$

Now we are in a position to prove

THEOREM 6.2: *The F-test, which rejects the hypothesis when $V$ is greater than some constant, determines a uniformly most powerful similar region for testing $\omega:\sigma_a^2 = 0$ against $\Omega - \omega:\sigma_a^2 > 0$.*

PROOF: We make use of the fact [7, p. 317] that if $T$ is a sufficient statistic for $\theta \varepsilon \omega$, and if $T$ is complete on $\omega$ then *all* similar tests of size $\alpha$,

$$E_\theta \varphi(W) \equiv \alpha, \quad \theta \varepsilon \omega, \quad W = (T, V)$$

have Neyman structure [12] with respect to $T$, i.e. satisfy

$$(6.16)\qquad \int \varphi(t, v) p_\theta(v \mid t)\, dv \underset{t}{\equiv} \alpha, \qquad \text{(a.e.) for all} \quad \theta \varepsilon \omega.$$

Subject to this we wish to maximize the power at a *particular* alternative $\theta_1 \varepsilon \Omega - \omega$; that is we desire

$$\int \left\{ \int \varphi(t, v) p_{\theta_1}(v \mid t)\, dv \right\} p_{\theta_1}(t)\, dt = \text{maximum}.$$

Using (6.14) and (6.15) these conditions become

$$(6.17)\qquad c \int_0^1 \varphi(t, v) v^{\frac{\nu_2-2}{2}} (1 - v)^{\frac{\nu_4-2}{2}}\, dv = \alpha$$

and

$$c \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} \left\{ \int_{0}^{1} \varphi(t, v) p_{\theta_1}(v \mid t)\, dv \right\} p_{\theta_1}(t)\, dt = \max$$

respectively, where

$$p_{\theta_1}(v \mid t) = \frac{u^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)-\frac{u}{2\lambda_4}}}{p_{\theta_1}(u)}.$$

This will be achieved if for each value of $t$ we have

$$(6.18) \quad c \int_0^1 \varphi(t, v) u^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)-\frac{u}{2\lambda_4}}\, dv = \max,$$

where we recall $t = (\bar{x}, s_3, s_6, u)$. But, finding for fixed $t$ (and *a fortiori* for fixed $u$) a test $\varphi(t, v)$ satisfying (6.17) and (6.18) is a problem whose solution is given at once by the fundamental Neyman-Pearson lemma to be $\varphi(t, v) = 1$ when

$$cu^{\frac{\nu_2+\nu_4-2}{2}} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}} e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)-\frac{u}{2\lambda_4}} > cv^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}}$$

or

$$e^{-\frac{uv}{2}\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_4}\right)} > c(\theta_1, t)$$

or

$$e^{kv} > c(\theta_1, t), \qquad k > 0.$$

or

$$v > c(\theta_1, t).$$

The "constant", $c = c(\theta_1, t)$ is determined by (6.17) or

$$\frac{1}{B\left(\frac{\nu_2}{2}, \frac{\nu_4}{2}\right)} \int_{v>c} v^{\frac{\nu_2-2}{2}} (1-v)^{\frac{\nu_4-2}{2}}\, dv = \alpha.$$

Consequently $c$ is independent of both $\theta_1$ and $t$,

$$\varphi(t, v) = 1 \quad \text{when} \quad v = \frac{s_2}{s_2 + s_4} > c$$

and the usual $F$-test is u.m.p. similar.[5]

Of course, Theorem 6.2 was proved only for the balanced two-way classification, but using the standard form in the next section it can be proved in the same way for the balanced one-way classification.

To show where the proof breaks down when applied to testing the hypothesis

[5] In commenting on an earlier draft of this paper, Dr. Werner Gautschi pointed out that in testing $\omega: \sigma_a^2 + \sigma_b^2 = 0$, the $T$ corresponding to (6.3) namely $T = (\bar{X}, S_2 + S_3 + S_4, S_5)$ is complete on $\omega$, but the method of Theorem 6.2 does not seem to help one show that the test based on $(S_2 + S_3)/(S_2 + S_3 + S_4)$ is u.m.p. similar.

of $\omega:\sigma_{ab}^2 = 0$ against $\Omega - \omega:\sigma_{ab}^2 > 0$ we try to prove the analogue of Theorem 6.1. The region $\Omega$ is still given by (6.1), but $\omega$ is now given by

$$\omega = \{\theta \mid -\infty < \mu < \infty; \quad 0 < \lambda_5 = \lambda_4 \leqq \lambda_2 \leqq \lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 < \infty; \\ \lambda_4 \leqq \lambda_3 \leqq \lambda_1 < \infty\}.$$

Clearly a sufficient statistic under $\omega$ is $T = (\bar{X}, S_2, S_3, U')$, where

$$U' = (S_4 + S_5).$$

Now

$$E_\theta f(T) = c(\lambda_2 + \lambda_3 - \lambda_4) \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} f(t) g(t, \lambda_2 + \lambda_3 - \lambda_4) h(t, \theta) \, du' \, ds_2 \, ds_3 \, d\bar{x}$$

where

$$g(t, \lambda_2 + \lambda_3 - \lambda_4) = \exp\left\{ -\frac{N\bar{x}^2}{2(\lambda_2 + \lambda_3 - \lambda_4)} \right\} s_2^{\frac{\nu_2 - 2}{2}} s_3^{\frac{\nu_3 - 2}{2}} (u')^{\frac{\nu_4 + \nu_5 - 2}{2}}$$

and

$$h(t, \theta) = \exp\left\{ \frac{N\mu\bar{x}}{\lambda_2 + \lambda_3 - \lambda_4} - \frac{s_2}{2\lambda_2} - \frac{s_3}{2\lambda_3} - \frac{u'}{2\lambda_4} \right\}.$$

The proof of Theorem 6.1 made use of the fact that

$$\exp\left\{ -\frac{N\bar{x}^2}{2\lambda_3} \right\} \exp\left\{ -\frac{s_3}{2\lambda_3} \right\} = \exp\left\{ -\frac{s_3^*}{2\lambda_3} \right\}$$

for $S_3^* = S_3 + N\bar{X}^2$. This method will not work here because the $\lambda_i$ associated with the mean, viz. $\lambda_1 \equiv \lambda_2 + \lambda_3 - \lambda_4$, does not equal any other $\lambda_i$. However, a lemma due to Gautschi[6] [17] and appearing in this issue of the *Annals* can be used to show completeness under this $\omega$ and thus that the $F$-test of $\sigma_{ab}^2 = 0$ is u.m.p. similar.

**7. Likelihood ratio test for the balanced one way classification.** We shall show that for the *balanced*, *one-way* classification the likelihood ratio (L.R.) test is not the $F$-test, but for purposes of significance testing we can act as if it were. Let us consider $I$ populations where the $j$th measurement on the $i$th population is given by

$$X_{ij} = \mu + e_i^A + e_{ij}, \tag{7.1}$$

$$i = 1, 2, \cdots, I; \quad j = 1, 2, \cdots, J; \quad IJ = N. \tag{7.2}$$

The usual Model II assumptions are made, namely, that $\mu$ is a constant and $e_i^A$, $e_{ij}$ are normally and independently distributed with zero means and vari-

[6] Dr. Gautschi independently derived the standard form which proved so useful in this work.

ances $\sigma_a^2$, $\sigma_e^2$ respectively. Let $D$ be the usual orthogonal transformation that transform $X_{ij}$, suitably ordered, to $Z_{ij}$ which have the standard distribution

$$(7.3) \qquad \frac{|\Lambda|^{-\frac{1}{2}}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}\left[\frac{s_1}{\lambda_1}+\frac{s_2}{\lambda_2}+\frac{s_3}{\lambda_3}\right]\right\},$$

where $S_1 = (Z_{11} - \sqrt{N}\mu)^2$, $S_2 = \sum_{i=2}^{I} Z_{i1}^2$, $S_3 = \sum_{j=2}^{J} \sum_{i=1}^{I} Z_{ij}^2$ and

$$(7.4) \qquad \lambda_1 = \lambda_2 = \sigma_e^2 + J\sigma_a^2$$

$$(7.5) \qquad \lambda_3 = \sigma_e^2$$

$$(7.6) \qquad |\Lambda| = \lambda_1\lambda_2^{I-1}\lambda_3^{I(J-1)}.$$

Clearly

$$(7.7) \qquad \lambda_2 \geqq \lambda_3 > 0.$$

To test $H_0:\sigma_a^2 = 0$ (or $\lambda_2 = \lambda_3$) it is well known that the usual $F$-test is equivalent to rejecting $H_0$ if $G > C$ where $G = S_2/S_3$ and $C$ is a constant which depends on the level of significance.

The maximum likelihood (M.L.) estimates, $\hat{\mu}_\Omega$, $\hat{\lambda}_{2\Omega}$, $\hat{\lambda}_{3\Omega}$, are values which maximize the likeiihood (7.3) subject to the condition that (7.7) is satisfied by the estimates, i.e.

$$(7.8) \qquad \hat{\lambda}_{2\Omega} \geqq \hat{\lambda}_{3\Omega}.$$

Equating to zero the derivatives of the likelihood with respect to $\mu$, $\lambda_2$, $\lambda_3$ one gets as solutions

$$(7.9) \qquad \hat{\mu}_\Omega = z_{11}/N$$

$$(7.10) \qquad \hat{\lambda}'_{2\Omega} = S_2/I$$

$$(7.11) \qquad \hat{\lambda}'_{3\Omega} = S_3/[I(J-1)].$$

Since differentiation *may* give, as solutions, values which do not satisfy condition (7.8), these estimates have primes to distinguish them from the "correct" M.L. estimates, which do satisfy (7.8) and are unprimed. That is, if

$$\hat{\lambda}'_{2\Omega} \geqq \hat{\lambda}'_{3\Omega},$$

then (7.10) and (7.11) are the correct M.L. estimates. Because $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$ is equivalent to $G < (J-1)^{-1}$ it remains only to see what the estimates are when (7.10) and (7.11) do not give the "correct" M.L. estimates, i.e., when $G < (J-1)^{-1}$. Since $L$, the logarithm of the likelihood may be written as a function of $\lambda_2$ plus a function of $\lambda_3$ it is clear that the values of $\lambda_2$ that maximize $L$, considered as a mathematical function defined for all positive $\lambda_2$ and $\lambda_3$, rather than as a likelihood (i.e. disregarding the restriction $\lambda_2 \geqq \lambda_3$), for fixed $\lambda_3$ is the same $\lambda_2$ as is given by (7.10) and similarly for the value of $\lambda_3$ that maximizes $L$ for fixed $\lambda_2$. Also $\partial L/\partial\lambda_2 \lesseqgtr 0$ or $\partial L/\partial\lambda_3 \lesseqgtr 0$ according as $\lambda_2 \gtrless S_2/I$ or $\lambda_3 \gtrless S_3/[I(J-1)]$. This means that for any fixed $\lambda_2$, $L$ decreases

as $\lambda_3$ moves away from $\hat{\lambda}'_{3\Omega}$ in either direction and similarly for $\lambda_2$ and $\hat{\lambda}'_{2\Omega}$ when $\lambda_3$ is fixed. Now, by (7.8), the point $(\hat{\lambda}_{2\Omega}, \hat{\lambda}_{3\Omega})$ in the $\lambda_2, \lambda_3$ plane cannot lie above the line $\lambda_3 = \lambda_2$. Suppose it were (strictly) below this line and $(\hat{\lambda}'_{2\Omega}, \hat{\lambda}'_{3\Omega})$ were above the line, i.e. $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$. If $\hat{\lambda}_{3\Omega} < \hat{\lambda}'_{3\Omega}$, then one can increase $L$ by increasing $\hat{\lambda}_{3\Omega}$; if $\hat{\lambda}_{3\Omega} \geqq \hat{\lambda}'_{3\Omega}$, $L$ can be increased by decreasing $\hat{\lambda}_{2\Omega}$. In both of these cases the assumption that $L$ is maximized at $(\hat{\lambda}_{2\Omega}, \hat{\lambda}_{3\Omega})$ is violated. Hence, whenever $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$, the "correct" maximum likelihood estimates are on the line, $\lambda_3 = \lambda_2$, which is the $\omega$ region. Thus maximum likelihood corrects negative estimates by making them zero. The maximum likelihood estimates are *then*

$$\hat{\mu}_\omega = \frac{Z_{11}}{N} \tag{7.12}$$

$$\hat{\lambda}_{2\omega} = \hat{\lambda}_{3\omega} = \frac{S_2 + S_3}{IJ}. \tag{7.13}$$

From (7.3) it can be seen that the square of the likelihood ratio is

$$R^2 = \frac{|\hat{\Lambda}_\Omega|}{|\hat{\Lambda}_\omega|} \exp\{z'\hat{\Lambda}_\Omega^{-1}z - z'\hat{\Lambda}_\omega^{-1}z\},$$

where $z' = \{z_{11} - \sqrt{N}\hat{\mu}, z_{21}, z_{31}, \cdots, z_{I1}; z_{12} \cdots z_{IJ}\}$. The subscripts are ordered as in (5.2) and $|_\Omega\hat{\Lambda}|$ and $|\hat{\Lambda}_\omega|$ are the maximum likelihood estimates of $|\Lambda|$. Since both $z'\hat{\Lambda}_\Omega^{-1}z$ and $z'\hat{\Lambda}_\omega^{-1}z$ can be shown to equal $N$, by a procedure given in the next section, $R^2 = |\hat{\Lambda}_\Omega| / |\hat{\Lambda}_\omega|$. By (7.6), this becomes

$$R^2 = \frac{\hat{\lambda}_{2\Omega}^I \hat{\lambda}_{3\Omega}^{I(J-1)}}{\hat{\lambda}_{3\omega}^{IJ}} \tag{7.14}$$

which is unity when $G < (J-1)^{-1} = G_0$, say. For $G \geqq G_0$, (7.9) to (7.13) imply that

$$R^2 = \frac{J^{IJ}}{(J-1)^{I(J-1)}} \frac{S_2^I S_3^{I(J-1)}}{(S_2 + S_3)^{IJ}}, \tag{7.15}$$

whence

$$R^{2/I} = KG\left(\frac{1}{G+1}\right)^J, \quad K = J^J/[(J-1)^{J-1}] > 0. \tag{7.16}$$

For values of $R$ below one and values of $G$ above $(J-1)^{-1}$, the L.R. test and the $G$ or $F$ test will now be shown to be equivalent. Since low values of $\lambda$ are significant, to show the equivalence of the two tests for this range of $G$ it is only necessary to show that $R^{2/I}$ is a decreasing function of $G$ or that

$$\frac{d}{dG}\left[G\left(\frac{1}{1+G}\right)^J\right] = \frac{1 + G - JG}{(1+G)^{J+1}} \tag{7.17}$$

is negative. Clearly, (7.17) is negative when $1 + G - JG < 0$ which is equivalent to $G > (J-1)^{-1} = G_0$. Also if $G = G_0$ in (7.16), $R = 1$. We have al-

ready seen that $R = 1$ when $G < G_0$. Now, let $\alpha_0 = \Pr\{G > G_0\}$, which is the probability that $R < 1$. Hence $1 - \alpha = \Pr\{R = 1\}$. Thus the atomic positive probability mass at $R = 1$ means that there are no L.R. tests of $\sigma_a^2 = 0$, for the balanced, one-way classification, with level of significance greater than $\alpha_0$ but less than 1. However, when an L.R. significance test does exist it is the $F$-test. Since $F = I(J - 1)G/(I - 1)$, $G > G_0$ is equivalent to $F > F_0$, where $F_0 = I/(I - 1)$. For all significance levels up to and including the 25 per cent level [9] the percentage points of $F$ with $(I - 1)$ and $I(J - 1)$ degrees of freedom for finite values of $I$ and $J$ greater than 1 are greater than $F_0$ while the 50 percentage points are less than $F_0$ for all these values of $I$ and $J$. Inasmuch as it is unlikely that one wishes to use a significance level between 25 and 50 per cent, for all *practical* purposes, the $F$-test and L.R. test are equivalent in the case of the *balanced-one-way* classification.

Although the $F$-tests of no population effect are the same under Models I and II, this quirk of the L.R. test does not exist in Model I. It is known that then the L.R. test is precisely the $F$-test. In Model I, the L.R. and the $F$-statistic are strictly decreasing functions of one another *and* there is no positive probability mass at $R = 1$.

It is of interest to note that there is a modified L.R. test which is equivalent to the $F$-test for the Model II, balanced, one-way classification. One can reason that if in (7.10) and (7.11) $\hat{\lambda}'_{2\Omega} < \hat{\lambda}'_{3\Omega}$, then the estimate of $\sigma_a^2$ as given by (7.4) and (7.5) is negative. Then one way to modify or "correct" the estimates so that the estimate of $\sigma_a^2$ is zero, is to use as estimates (although they are no longer M.L.),

$$\hat{\lambda}_{2\Omega c} = \hat{\lambda}_{3\Omega c} = \frac{S_3}{I(J - 1)}. \tag{7.18}$$

If these are put in (7.14), and if $K$ is a positive constant

$$R^2 = \left(\frac{K}{1 + G}\right)^{IJ} \tag{7.19}$$

which is a strictly decreasing function of $G$. Hence, if the estimates given by (7.18) are used when $\hat{\lambda}'_{2\Omega} > \hat{\lambda}'_{3\Omega}$, this modified L.R. test is equivalent to the $F$-test. Little can be said for this procedure, since the information in $S_2$ is not used and when the estimate of $\sigma_a^2$ is negative one can argue almost as easily, by ignoring the information in $S_3$, that the corrected estimates should be

$$\hat{\lambda}_{2\Omega c} = \hat{\lambda}_{3\Omega c} = S_2/I. \tag{7.20}$$

If these are used in (7.14),

$$R^2 = \left(\frac{K}{1 + 1/G}\right)^{IJ} \tag{7.21}$$

where $K$ is again a positive constant. This is a strictly *increasing* function of $G$

and the modified test using it is certainly not equivalent to an $F$-test with large values significant.

**8. Likelihood ratio test for the balanced two-way classification.** This section will be devoted to showing by means of a counter-example that when testing $\omega: \sigma_a^2 = 0$ in the two-way classification, not only is the L.R. test not the $F$-test, but (unlike the balanced one-way classification) is not even equivalent to it for small levels of significance. The L.R. test is a function of $S_2$, $S_3$ and $S_4$, while the $F$-test is a function of only $S_2$ and $S_4$. From (5.5) the logarithm of the likelihood for $\theta \varepsilon \Omega$, $\Omega$ given by (6.1), is

$$L_\theta = -\frac{N}{2}\log 2\pi - \frac{1}{2}\left\{\log|\Lambda| + \sum_{i=1}^{5}\frac{s_i}{\lambda_i}\right\} \tag{8.1}$$

where $|\Lambda| = \lambda_1\lambda_2^{\nu_2}\lambda_3^{\nu_3}\lambda_4^{\nu_4}\lambda_5^{\nu_5}$. The $s_i$ are defined by (5.6) and the $\lambda$'s by (3.3). Recall that for all $\theta \varepsilon \Omega$

$$\lambda_1 = \lambda_2 + \lambda_3 - \lambda_4 \tag{8.2}$$

and

$$\lambda_1 \geqq \lambda_2 \geqq \lambda_4\,; \qquad \lambda_1 \geqq \lambda_3 \geqq \lambda_4\,; \qquad \lambda_4 \geqq \lambda_5 > 0. \tag{8.3}$$

Rather than maximize $L_\theta$ subject to (8.2) we shall use a more general side condition, use of which will be made below, namely to maximize $L_\theta$ subject to $\sum_1^5 b_i\lambda_i = 0$ and $\sum_1^5 c_i\lambda_i = 0$ by making use of Lagrange multipliers $\beta/2$ and $\gamma/2$. Let

$$M = L_\theta + \frac{\beta}{2}\sum_{i=1}^{5} b_i\lambda_i + \frac{\gamma}{2}\sum_{i=1}^{5} c_i\lambda_i.$$

Equating to zero the derivative of $M$ with respect to $\beta$, $\gamma$, $\mu$, $\lambda_i$, $(i = 1, 2, \cdots, 5)$ one obtains

$$\hat{\mu} = Z_{111}/\sqrt{N} \tag{8.4}$$

$$-\nu_i + \frac{S_i}{\hat{\lambda}_i} + \beta b_i\hat{\lambda}_i + \gamma c_i\hat{\lambda}_i = 0, \qquad i = 1, 2, \cdots, 5 \tag{8.5}$$

$$\sum_1^5 b_i\hat{\lambda}_i = 0, \qquad \sum_1^5 c_i\hat{\lambda}_i = 0, \tag{8.6}$$

where the carats indicate that these are the maximizing values. Adding the five equations in (8.5) and making use of (8.6) we obtain

$$\sum_{i=1}^{5}\frac{S_i}{\hat{\lambda}_i} = \sum_{i=1}^{5}\nu_i = N,$$

and the exponent in (5.5) is $-N/2$ when the maximizing values of the parameters are used. Thus the well-known result [18] when there is no condition on the $\lambda$'s is also true if the $\lambda$'s are linearly dependent.

Under $\Omega$, $b_1 = b_4 = 1$, $b_2 = b_3 = -1$, $b_5 = 0$, $c_i = 0$, $(i = 1, 2, \cdots, 5)$ and (8.4)–(8.6) become

$$\hat{\mu}_\Omega = Z_{111}/\sqrt{N} \tag{8.7}$$

$$\begin{aligned} \hat{\lambda}_{1\Omega}^{-1} - \beta &= 0 \\ \nu_2\hat{\lambda}_{2\Omega}^{-1} - S_2\hat{\lambda}_{2\Omega}^{-2} + \beta &= 0 \\ \nu_3\hat{\lambda}_{3\Omega}^{-1} - S_3\hat{\lambda}_{3\Omega}^{-2} + \beta &= 0 \\ \nu_4\hat{\lambda}_{4\Omega}^{-1} - S_4\hat{\lambda}_{4\Omega}^{-2} - \beta &= 0 \end{aligned} \tag{8.8}$$

$$\hat{\lambda}_{5\Omega} = S_5/\nu_5 . \tag{8.9}$$

If the solutions of (8.2) and (8.8) satisfy (8.3), they are also M.L. estimates. If not then one would have to get the "correct" M.L. estimates by some procedure similar to the one used in the previous section. This will be unnecessary because we shall show that even when these solutions satisfy (8.3) the L.R. statistic is not a function of $F$ alone. Hereafter we confine ourselves to the part of the $z$ space where (8.3) is satisfied by the stationary values. Eliminating the Lagrange multiplier, and performing some simplifications one may write (8.2) and (8.8) as

$$\begin{aligned} \hat{\lambda}_{1\Omega} &= \hat{\lambda}_{2\Omega} + \hat{\lambda}_{3\Omega} - \hat{\lambda}_{4\Omega} \\ S_2\hat{\lambda}_{2\Omega}^{-1} &= \nu_2 + \hat{\lambda}_{2\Omega}\hat{\lambda}_{1\Omega}^{-1} \\ S_3\hat{\lambda}_{3\Omega}^{-1} &= \nu_3 + \hat{\lambda}_{3\Omega}\hat{\lambda}_{1\Omega}^{-1} \\ S_4\hat{\lambda}_{4\Omega}^{-1} &= \nu_4 - \hat{\lambda}_{4\Omega}\hat{\lambda}_{1\Omega}^{-1} \end{aligned} \tag{8.10}$$

Similarly, the logarithm of the likelihood under $\omega:\sigma_a^2 = 0$ or $\lambda_1 = \lambda_3$, $\lambda_2 = \lambda_4$ is given by (8.1) subject to

$$b_1 = b_4 = 1, \qquad b_2 = b_3 = -1, \qquad b_5 = 0$$

$$c_2 = 1, \qquad c_4 = -1, \qquad c_1 = c_3 = c_5 = 0.$$

Using this last condition (8.4)–(8.6) can be simplified to

$$\begin{aligned} \hat{\mu}_\omega &= Z_{111}/\sqrt{N} \\ \hat{\lambda}_{3\omega} &= S_3/I \\ \hat{\lambda}_{4\omega} &= (S_2 + S_4)/I(J-1) \\ \hat{\lambda}_{5\omega} &= S_5/IJ(K-1). \end{aligned} \tag{8.11}$$

As in the estimate under $\Omega$ we treated only the case when the stationary values satisfy $\hat{\lambda}_{3\omega} \geqq \hat{\lambda}_{4\omega} \geqq \hat{\lambda}_{5\omega} > 0$.

Since the exponent of the likelihood when the estimates under $\Omega$ *or* $\omega$ are

inserted has been shown above to equal $-N/2$, the square of the likelihood ratio is

$$R^2 = \frac{|\hat{\Lambda}_\Omega|}{|\hat{\Lambda}_\omega|} = \frac{L'_\Omega}{L'_\omega} \tag{8.12}$$

where

$$L'_\Omega = (\hat{\lambda}_{2\Omega} + \hat{\lambda}_{3\Omega} - \hat{\lambda}_{4\Omega})\hat{\lambda}_{2\Omega}^{\nu_2}\hat{\lambda}_{3\Omega}^{\nu_3}\hat{\lambda}_{4\Omega}^{\nu_4} \tag{8.13}$$

$$L'_\omega = (S_3/I)^I[(S_2 + S_4)/I\nu_2]^{I\nu_2} \tag{8.14}$$

and $\hat{\lambda}_{i\Omega}$, $i = 2, 3, 4$ satisfy (8.10). We have seen that the $F$-test of $\omega:\sigma_a^2 = 0$ is a function of $S_2$ and $S_4$ alone and does not depend on $S_3$. It appears that $R^2$ *may* depend on $S_3$ since its denominator, $L'_\omega$ does. However $L'_\Omega$ *may* equal $S_3^I$ times a factor independent of $S_3$, in which case $R^2$ will be independent of $S_3$. It was shown [3], by comparing the solutions, (8.12) for two examples differing only in values for $s_3$, that $R^2$ *does* depend on $S_3$.

In Section 10 it will be shown that both the L.R. and $F$-tests are invariant tests, but the $F$-test is to be preferred since it has an optimum property, namely, of being the u.m.p. similar test.

**9. Uniformly most powerful invariant test in the balanced, one-way classification is the $F$-test.** It will be shown that for the balanced, one-way classification, when the standard variable $Z$ has distribution (7.3) the u.m.p. invariant test of $\omega:\sigma_a^2 = 0$ against $\Omega - \omega:\sigma_a^2 > 0$ (or using (7.4) and (7.5) of $\omega:\theta = 1$ against $\Omega - \omega:\theta > 1$ where $\theta = \lambda_2/\lambda_3$) is the $F$-test. We partition the $Z$ vector as follows. Let $Z' = [Z_{(1)}, Z'_{(2)}, Z'_{(3)}]$ where $Z_{(1)} = Z_{11}$, $Z_{(2)}$ is the column vector whose elements are $Z_{i1}$, $i = 2, 3, \cdots, I$ and $Z_{(3)}$ is the column vector whose elements are $Z_{ij}$; $i = 1, 2, \cdots, I$; $j = 2, 3, \cdots, J$. The elements of $Z_{(2)}$ and $Z_{(3)}$ may be ordered in any way. Clearly the problem remains invariant under the following groups of transformations, each of which is a normal subgroup of the product group of the previous ones:

$$Z^*_{(1)} = Z_{(1)} + c, \qquad Z^*_{(\alpha)} = Z_{(\alpha)}, \quad \alpha = 2, 3. \tag{9.1}$$

$$Z^*_{(\alpha)} = D_{(\alpha)}Z_{(\alpha)}; \qquad D_{(\alpha)} \text{ orthogonal}, \quad \alpha = 1, 2, 3 \tag{9.2}$$

$$Z^*_{(\alpha)} = cZ_{(\alpha)}; \qquad \alpha = 1, 2, 3; \ c \neq 0. \tag{9.3}$$

A maximal invariant [8] under the product of all three groups is

$$G = \left(\textstyle\sum_{i=2}^I Z_{i1}^2\right)/\left(\textstyle\sum_{j=2}^J \sum_{i=1}^I Z_{ij}^2\right) = \frac{S_2}{S_3}.$$

It may be pointed out that unlike Model I the group of orthogonal transformations is unnecessary if we agree to base all decisions on the *sufficient* statistic $(Z_{11}, S_2, S_3)$ of Section 7. Starting with this statistic the first and third group of transformations (additive and multiplicative group) will lead to $G$ as a maximal invariant in the class of sufficient statistics.

To show that the test which determines the critical region $G > c$ (or the equivalent $F$-test which rejects $\omega$ when $W = \nu_3 G/\nu_2 > c$) is the u.m.p. invariant test one need only show it is the u.m.p. test based on $G$. Under $\omega$, $W$ is distributed as $F$ with $\nu_2$ and $\nu_3$ degrees of freedom, while, under $\Omega - \omega$, it is distributed as $\theta$ times $F$ with $\nu_2$ and $\nu_3$ degrees of freedom, i.e. the probability density of $G$ is

$$p_\theta(g) = c\theta^{\frac{\nu_3}{2}} g^{\frac{\nu_2-2}{2}} (\theta + g)^{-\left(\frac{\nu_2+\nu_3}{2}\right)}, \qquad \theta \geqq 1, \tag{9.4}$$

where $\theta = \lambda_2/\lambda_3$. By the Neyman-Pearson lemma the most powerful test of $\theta = 1$ based on $G$ against a particular alternative $\theta = \theta_0 > 1$ is given by $\varphi(g) = 1$ when

$$\theta_0^{\frac{\nu_3}{2}} \left(\frac{1 + g}{\theta_0 + g}\right)^{\frac{\nu_2+\nu_3}{2}} > c$$

or

$$\frac{1 + g}{\theta_0 + g} > c. \tag{9.5}$$

The left member of (9.5) is an increasing function of $g$, since its derivative with respect to $g$ is $(\theta_0 - 1)/(\theta_0 + g)^2$ which is positive. Hence this test is equivalent to $\varphi(g) = 1$ when $g > c$. Since the value of $c$ is determined by integrating the upper tail of (9.4) for $\theta = 1$, it is not dependent on the particular alternative. Thus for the balanced one-way classification one may replace the class of similar tests by the somewhat more reasonable class of invariant tests and show that in this more reasonable class the usual $F$-test of $\sigma_a^2 = 0$ against $\sigma_a^2 > 0$ is also u.m.p.

**10. Invariance in the balanced two-way classification.** It will now be shown why there *may* not be *any* uniformly most powerful invariant test in the case of the balanced two-way classification. We are interested in the test of $\omega:\sigma_a^2 = 0$ against $\Omega - \omega:\sigma_a^2 > 0$ (or, if we let $\psi_1 = \lambda_2/\lambda_4$, of testing $\omega:\psi_1 = 1$ against $\Omega - \omega:\psi_1 > 1$) for the standard variate $Z$ whose distribution is given by (5.5). The group of transformations analogous to those in the last section will be considered. As in that section we partition the $Z$ vector thus:

$$Z' = [Z_{(1)}, Z'_{(2)}, Z'_{(3)}, Z'_{(4)}, Z'_{(5)}],$$

where $Z_{(1)} = Z_{111}$ and $Z_{(\alpha)}$, for $\alpha = 2, 3, 4, 5$, is the column vector of the $Z$'s (in any order) appearing in the sums $S_\alpha$ of (5.6). The problem remains invariant under the same types of groups of transformations as in the preceding section, namely (9.1) for $\alpha = 2, 3, 4, 5$ and (9.2), (9.3) for $\alpha = 1, \cdots, 5$. A maximal invariant under the product group of the three groups, is $U$, $V$, $W$ where

$$U = S_2/S_4, \quad V = S_3/S_4, \quad W = S_4/S_5. \tag{10.1}$$

Any test based on $U$, $V$, $W$ will have power based on the maximal invariant induced in the parameter space, namely,

$$\psi = (\psi_1, \psi_2, \psi_3), \tag{10.2}$$

where

$$(10.3) \qquad \psi_1 = \lambda_2/\lambda_4, \qquad \psi_2 = \lambda_3/\lambda_4, \qquad \psi_3 = \lambda_4/\lambda_5.$$

As in the balanced one-way classification (Section 9), the orthogonal group of transformations corresponding to (9.2) is unnecessary if we agree to base all decisions on the sufficient statistic $(Z_{111}, S_2, S_3, S_4, S_5)$ of Section 5.

By transforming the density of $(S_2, S_3, S_4, S_5)$ as given in (5.9) to that of $(S_5, U, V, W)$ and integrating out $s_5$ we obtain [3]

$$(10.4) \qquad p_\psi(u, v, w) = \frac{\Gamma\left(\frac{\nu_2 + \nu_3 + \nu_4 + \nu_5}{2}\right)}{\psi_1^{\frac{\nu_2}{2}} \psi_2^{\frac{\nu_3}{2}} \psi_3^{\frac{\nu_2+\nu_3+\nu_4}{2}}} \frac{u^{\frac{\nu_2-2}{2}} v^{\frac{\nu_3-2}{2}} w^{\frac{\nu_2+\nu_3+\nu_4-2}{2}}}{\delta^{\frac{\nu_2+\nu_3+\nu_4+\nu_5}{2}}}$$

where

$$(10.5) \qquad \delta = \delta(u, v, w; \psi) = \frac{uw}{\psi_1\psi_3} + \frac{w}{\psi_3} + \frac{vw}{\psi_2\psi_3} + 1$$

This shows that the density of $u$, $v$, $w$ is indeed dependent on $\lambda$ only through the maximal invariant $\psi = (\psi_1, \psi_2, \psi_3)$. The Neyman-Pearson lemma gives as the most powerful test of $H_0:\lambda = \lambda^0$ against $H_1:\lambda = \lambda^1$ (where $\lambda^i = (\lambda_2^i, \lambda_3^i, \lambda_4^i, \lambda_5^i)$, $\lambda_2^i/\lambda_4^i = \psi_1^i$, $i = 0, 1$ and $\psi_1^0 = 1 < \psi_1^1$), based on $(U, V, W)$ the one which rejects $H_0$ when

$$(10.6) \qquad \frac{p_{\psi^1}(u, v, w)}{p_{\psi^0}(u, v, w)} = \left[\frac{\frac{w(u+1)}{\psi_3^0} + \frac{vw}{\psi_2^0\psi_3^0} + 1}{\frac{uw}{\psi_1^1\psi_3^1} + \frac{w}{\psi_3^1} + \frac{vw}{\psi_2^1\psi_3^1} + 1}\right]^{\frac{\nu_2+\nu_3+\nu_4+\nu_5}{2}} \cdot \left(\frac{\psi_1^0}{\psi_1^1}\right)^{\frac{\nu_2}{2}} \left(\frac{\psi_2^0}{\psi_2^1}\right)^{\frac{\nu_3}{2}} \left(\frac{\psi_3^0}{\psi_3^1}\right)^{\frac{\nu_2+\nu_3+\nu_4}{2}} > c.$$

Since the distribution of $V$ and $W$ depend on $\psi_2^0$ and $\psi_3^0$ under $H_0$ there seems to be little likelihood of obtaining a *uniformly* most powerful invariant test based on a statistic involving $U$, $V$ and $W$ from (10.6). It was not obvious from the fact that the maximal invariant was vector valued that no u.m.p. invariant test exists, since conceivably (10.6) might involve only one of the elements of the vector. For example if (10.6) were a function of $U$ alone then once again the usual $F$-test would be uniformly most powerful. Although our probability ratio, (10.6) showed that there is no u.m.p. invariant test based on the given product group of transformations, there still may be one with respect to a larger group of transformations. For example, if in the last section we had stopped after the second group of transformations obtaining as a maximal invariant $S_2$, $S_3$ (rather than $S_2/S_3$) a situation analogous to (10.6) would have resulted. This may mean that another group of transformations, unknown to the author, may leave the problems invariant in the case of the balanced two-

way classification and the maximal invariant under the product of the four groups is $U$.

Even if there are no further invariant transformations, an optimum test in this case can be obtained by *decreasing* the class of invariant tests. We have seen that a maximal invariant under $G$ is the vector consisting of any three independent ratios of $S_2$, $S_3$, $S_4$, $S_5$ and that $G$ induced a group $\bar{G}$ under which a maximal invariant in the parameter space is the vector composed of the corresponding three ratios of $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$. But $\psi_1 = \lambda_2/\lambda_4$ (or its reciprocal) seems to be the only one that is independent of nuisance parameters under $\omega$. Also $S_2/S_4$ (or its reciprocal) appears to be the only part of the maximal invariant under $G$ whose distribution is a function of $\psi_1$ only. Thus, *it seems reasonable* to restrict our class to $S_2/S_4$. Then we obtain a u.m.p. test as in the last section. We now show

THEOREM 10.1: *Of all invariant tests of* $\omega$: $\sigma_a^2 = 0$ *against* $\Omega - \omega: \sigma_a^2 > 0$ *in the balanded two-way classification whose power is a function of* $\psi_1$ *only, the usual F-test is most powerful.*

PROOF: *If* it can be shown that $S_2/S_4$ is the only invariant statistic whose power is a function of $\psi$ only, the above assertion is true. However we have already shown a stronger result in Section 6, which includes this result, namely, the usual $F$-test is the u.m.p. similar test. Similarity in this example means

$$E_\theta\varphi(X) = \alpha, \qquad \theta \varepsilon \omega \text{ (i.e. } \psi_1 = 1),$$

while we want our test to satisfy

$$\begin{aligned} E_\psi\varphi(X) &= \text{const} = \alpha, \text{ say} \quad && \text{for } \psi \varepsilon \omega \text{ (i.e. } \psi_1 = 1) \\ &= f(\psi_1) && \psi \varepsilon \Omega - \omega \text{ (i.e. } \psi_1 > 1) \\ X &= h(U, V, W). \end{aligned}$$

By $\psi \varepsilon \omega$ we mean that the components of $\psi$ satisfy (6.2). There clearly is a similar test for every invariant test which is a function of $\psi_1$ only. Since the u.m.p. similar test is based on $U$, an invariant statistic, Theorem 10.1 is proved. Of course invariance added nothing in this case.

**11. Balanced multi-way classifications.** The procedure of Section 5 can be used to obtain the standard form for any balanced multi-way classification. The evaluation of $| \mathfrak{T} - \lambda \mathfrak{g} |$ just becomes a little more tedious as the number of factors increases. Of special interest is the case of the multi-fold, hierarchical or nested classification [6] model which is very useful in survey sampling theory [1]. The three-fold classification may be represented by

$$X_{ijkm} = \mu + e_i^A + e_{ij}^{AB} + e_{ijk}^{ABC} + e_{ijkm}$$

with $\mu$ a constant, $e_i^A$, $e_{ij}^{AB}$, $e_{ijk}^{ABC}$, $e_{ijkm}$, normally and independently distributed with means zero and variances $\sigma_a^2$, $\sigma_{ab}^2$, $\sigma_{abc}^2$, $\sigma_e^2$ and the range of subscripts as usual. In this special case the hypothesis that any variance component, except $\sigma_e^2$, equals zero can be tested by an $F$-test and the method of Section 6 can be

used to show these tests are u.m.p. similar. Even in this special case the methods of Section 9 cannot be used to show u.m.p. invariance unless the multi-fold classification is one-fold, which is the same as the one-way case treated in Section 9. However, Gautschi's [17] lemma must be used to prove that the usual $F$-tests are u.m.p. similar in the *non-hierarchical* multi-way classifications, when there are more than two classifications.

**12. Acknowledgements.** The author is indebted to many people, especially to Drs. H. Scheffé, H. Raiffa, T. W. Anderson and Erich Lehmann for helpful suggestions and to Dr. H. Scheffé for suggesting the problem that led to the paper.

## REFERENCES

[1] W. G. Cochran, *Sampling Techniques*, John Wiley and Sons, New York, 1953.
[2] S. L. Crump, "The estimation of components of variance in multiple classifications", unpublished Ph.D. thesis, Ames, Iowa, Iowa State College Library, 1947.
[3] L. H. Herbach, "Optimum properties of analysis of variance tests based on Model II and some generalizations of Model II", Scientific Paper No. 6, Engineering Statistics Laboratory, New York University, 1957 (Ph.D. Dissertation, Columbia University).
[4] P. L. Hsu, "Analysis of variance from the power function standpoint", *Biometrika*, Vol. 32 (1941), p. 62.
[5] P. L. Hsu, "On the power function of the $E^2$ test and the $T^2$ test", *Ann. Math. Stat.*, Vol. 16 (1945), p. 278.
[6] O. Kempthorne, *The Design and Analysis of Experiments*, John Wiley and Sons, New York, 1952.
[7] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation, Part I", *Sankhya*, Vol. 10 (1950), pp. 305–340.
[8] E. L. Lehmann, "Some principles of the theory of testing hypotheses", *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[9] M. Merrington and C. M. Thompson, "Tables of percentage points of the inverted beta (F) distribution", *Biometrika*, Vol. 33 (1943), pp. 73–88.
[10] A. M. Mood, *Introduction to the Theory of Statistics*, McGraw Hill, New York, 1950.
[11] J. A. Nelder, "The interpretation of negative components of variance", *Biometrika*, Vol. 41 (1954), pp. 544–548.
[12] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability", *Phila. Trans. Roy. Soc. London, Ser. A*, Vol. 236 (1937), pp. 333–380.
[13] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use", *Stat. Res. Memoirs*, Vol. 2 (1938), p. 126.
[14] A. Wald, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), p. 434.
[15] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, New Jersey, 1947.
[16] J. Wolfowitz, "The power of the classical tests associated with the normal distribution," *Ann. Math. Stat.*, Vol. 20 (1949), p. 540.
[17] W. Gautschi, "Some remarks on Herbach's paper, 'Optimum nature of the F-test for Model II in the balanced case,'" *Ann. Math. Stat.*, Vol. 30 (1959), pp. 960–963.
[18] T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, John Wiley, New York, 1958, p. 47.

# SOME REMARKS ON HERBACH'S PAPER, "OPTIMUM NATURE OF THE F-TEST FOR MODEL II IN THE BALANCED CASE"[1]

By Werner Gautschi †

*Indiana University*

**1. Summary.** The purpose of this note is to present a lemma which will settle a question of completeness left open in Section 6 of the above mentioned paper [5]. We give two applications of the lemma,

(i) by proving that, in addition to Herbach's results, also the standard $F$-test for $\sigma^2_{ab} = 0$ is a uniformly most powerful similar test,

(ii) by pointing out that the standard form introduced in [5] together with our lemma provide convenient tools to prove that in a balanced model II design (with the usual normality assumptions) *the standard estimates of variance components are minimum variance unbiased.* This result is well known ([2], [3]) and it has in fact been pointed out by Graybill and Wortham [3] that a completeness argument may be used to demonstrate the minimum variance property of the usual estimators for the variance components. The present lemma shows that the estimators do indeed have the necessary completeness property. We will follow Herbach's notation throughout.

**2. A completeness lemma.** The following lemma guarantees completeness for a certain class of probability densities to which the results of Lehmann and Scheffé do not apply directly. It takes care of a difficulty mentioned in [5], Section 6, which is caused when $g(\theta)$ does not equal one of the $\theta_i$ $(i = 2, \cdots, r)$. If $g(\theta)$ does, the product-densities could immediately be reduced to the exponential form considered by Lehmann and Scheffé in [7], Theorem 7.3. Our lemma is more general than the Lehmann and Scheffé Theorem 7.1 [7] in the sense that we allow instead of their $g''_{\theta''}(x'')$ to have $g''_{\theta',\theta''}(x'')$ which, however, we assume to factor into $h'_{\theta'}(x'')h''_{\theta''}(x'')$ with $h'_{\theta'}(x'') > 0$ and $\{h_{\theta''}(x'')\, d\mu^{x''}\}$ strongly complete. It is of course more special in that we take both $\mu^{x''}$ and $\mu^{x'|x''}$ as Lebesgue measure and for $g'_{\theta'}(x')$, $g''_{\theta',\theta''}(x'')$ specific functions. Our proof is modelled along the same lines as the one given by Lehmann and Scheffé in [7] p. 221.

Lemma: *Let*

$$\mathfrak{P}^t = \{P^t_\theta\,;\, \theta\ \varepsilon\ \mathfrak{D}\}, \qquad t = (t_2, \cdots, t_r), \theta = (\theta_2, \cdots, \theta_r)$$

$$\mathfrak{P}^{t_1} = \{P^{t_1}_{\theta_1,\theta}\,;\, (\theta_1, \theta)\ \varepsilon\ \mathfrak{D}_1 \times \mathfrak{D}\}, \qquad \theta_1 \text{ real}$$

Received October 8, 1957; revised January 27, 1959.

[1] This is a cut-down version of a paper in which the author independently considered standard forms for model II designs. He acknowledges, however, the priority of Dr. Herbach's approach (see [4] as compared to [1]) and restricts himself to giving some results supplementing those of Herbach.

† Werner Gautschi died on October 3, 1959. *Editor.*

*be two families of probability measures on the Borel sets of the Euclidean space $E_{r-1}$ and the real line $E_1$ respectively, having the densities*

$$p_\theta(t) = c(\theta)h(t_2, \cdots, t_r)e^{\theta_2 t_2 + \cdots + \theta_r t_r} \tag{1}$$

$$p_{\theta_1,\theta}(t_1) = c(\theta_1, \theta)e^{\rho(\theta)t_1^2 + \theta_1 t_1} \tag{2}$$

*with respect to Lebesgue measure. If $\mathfrak{D}_1$ is the real line and $\mathfrak{D}$ a Borel set in $E_{r-1}$ containing a non-degenerate $(r-1)$-dimensional interval then the family of product measures $\mathfrak{P} = \{P_{\theta_1,\theta}^{t_1} \times P_\theta^t ; (\theta_1, \theta) \varepsilon \mathfrak{D}_1 \times \mathfrak{D}\}$ is strongly complete (in the sense of Lehmann and Scheffé [7]).*

PROOF: Suppose

$$I = \iint f(t_1, t)p_{\theta_1,\theta}(t_1)p_\theta(t)\, dt_1\, dt = 0 \qquad (\text{a.e. } L^{\theta_1\times\theta}).^2 \tag{3}$$

Let $N$ be the set of parameter points $(\theta_1, \theta)$ for which $I \neq 0$. If $N_\theta$ denotes the $\theta$-section of $N$, i.e. $N_\theta = \{\theta_1 ; (\theta_1, \theta) \varepsilon N\}$, then $L^{\theta_1}(N_\theta) = 0$ except possibly for $\theta \varepsilon N_0$, where $L^\theta(N_0) = 0$.

According to Fubini's theorem we may write

$$I = \int p_{\theta_1,\theta}(t_1)\Phi(t_1, \theta)\, dt_1,$$

where $\Phi(t_1, \theta) = \int f(t_1, t)p_\theta(t)\, dt$. Since $p_{\theta_1,\theta}(t_1) > 0$, for fixed $\theta \not\varepsilon N_0$ the exceptional set of points $t_1$ for which the integral defining $\Phi(t_1, \theta)$ does not exist has $L^{t_1}$-measure zero. Furthermore, if $\theta \not\varepsilon N_0$, we can, in virtue of (2), rewrite (3) as

$$\int e^{\theta_1 t_1}\left[e^{\rho(\theta)t_1^2}\Phi(t_1, \theta)\right] dt_1 = 0 \qquad (\text{a.e. } L^{\theta_1}), \theta \not\varepsilon N_0.$$

From the unicity property of the bilateral Laplace transform (see, for instance, [8], Ch. VI, Theorem 6b) it follows that

$$\Phi(t_1, \theta) = 0 \qquad (\text{a.e. } L^{t_1}), \theta \not\varepsilon N_0.$$

Thus, if $S$ denotes the (measurable) set of points $(t_1, \theta)$ for which $\Phi$ is either not defined or $\neq 0$, almost every $\theta$-section of $S$ has $L^{t_1}$-measure zero, hence $L^{t_1\times\theta}(S) = 0$.

This in turn implies that almost all $t_1$-sections of $S$ have $L^\theta$-measure zero, i.e.

$$\Phi(t_1, \theta) = \int f(t_1, t)p_\theta(t)\, dt = 0 \qquad (\text{a.e. } L^\theta) \quad \text{if} \quad t_1 \not\varepsilon N_1,$$

where $L^{t_1}(N_1) = 0$. Since the family of probability densities $p_\theta(t)$ is strongly complete (Lehmann and Scheffé [7], Theorem 7.3) we conclude

---

[2] $L$ with a superscript denotes Lebesgue measure. The superscript indicates the space on which the measure is taken.

$$f(t_1, t) = 0 \qquad (\text{a.e. } \mathfrak{P}'), t_1 \varepsilon N_1,$$

from which $f(t_1, t) = 0$ (a.e. $\mathfrak{P}$) follows immediately.

**3. Applications.** (a) *Tests of hypotheses in balanced model* II *designs.* Consider the balanced two-way classification ([5], Section 6) and the hypothesis $\omega:\sigma_{ab}^2 = 0$. The statistic

$$T_1 = Z_{111}, \qquad T_2 = S_2, \qquad T_3 = S_3, \qquad T_4 = S_4 + S_5$$

is not only sufficient under $\omega$ but also complete on $\omega$. In fact, if we let

$$\theta_1 = \frac{\sqrt{N}\mu}{\lambda_2 + \lambda_3 - \lambda_4}, \qquad \theta_2 = -\frac{1}{2\lambda_2}, \qquad \theta_3 = -\frac{1}{2\lambda_3}, \qquad \theta_4 = -\frac{1}{2\lambda_4},$$

the densities of $T_1$ and $T = (T_2, T_3, T_4)$ are easily recognized to have the form given in our lemma. Proceeding therefore in the same fashion as in [5], Section 6, we would find that *also the standard F-test of the hypothesis* $\omega:\sigma_{ab}^2 = 0$ *is a uniformly most powerful similar test.* The same situation prevails in higher order classifications. As is well known, in a complete $n$-way classification $F$-tests exist for the non-existence of anyone of the $(n - 1)$st or $(n - 2)$nd order interactions. All these tests are uniformly most powerful similar tests.

(b) *Point estimation in balanced model* II *designs.* To fix the ideas consider the standard form for the balanced two-way classification. A sufficient statistic for the parameters involved is

$$T_1 = Z_{111}, \qquad T_2 = S_2, \cdots, \qquad T_5 = S_5. \tag{4}$$

If we let

$$\theta_1 = \frac{\sqrt{N}\mu}{\lambda_2 + \lambda_3 - \lambda_4}, \qquad \theta_2 = -\frac{1}{2\lambda_2}, \cdots, \qquad \theta_5 = -\frac{1}{2\lambda_5},$$

the densities of $T_1$ and $T = (T_2, \cdots, T_5)$ are again of the form given in our lemma and thus the statistic (4) is complete on $\Omega$. Unbiased estimates for the variance components, in terms of (4), are

$$\hat{\sigma}_e^2 = \frac{T_5}{\nu_e}, \quad \hat{\sigma}_{ab}^2 = \frac{1}{K}\left[\frac{T_4}{\nu_{ab}} - \frac{T_5}{\nu_e}\right], \quad \hat{\sigma}_b^2 = \frac{1}{IK}\left[\frac{T_3}{\nu_b} - \frac{T_4}{\nu_{ab}}\right], \quad \hat{\sigma}_a^2 = \frac{1}{JK}\left[\frac{T_2}{\nu_a} - \frac{T_4}{\nu_{ab}}\right], \tag{5}$$

where $\nu_a = I - 1$, $\nu_b = J - 1$, $\nu_{ab} = (I - 1)(J - 1)$, $\nu_e = IJ(K - 1)$ and are therefore minimum variance unbiased estimates ([6], Theorem 5.1). On the other hand the standard estimates in terms of the various mean squares have the same distribution as those in (5) and must consequently be of minimum variance among all unbiased estimates based on the original observation vector $X$.

Higher order layouts could be treated in a similar manner.

## REFERENCES

[1] W. GAUTSCHI, "On an optimal property of variance-components estimates" (Abstract), *Ann. Math. Stat.* Vol. 28 (1957), p. 1058.

[2] F. A. GRAYBILL, "On quadratic estimates of variance components", *Ann. Math. Stat.* Vol. 25 (1954), pp. 367–372.

[3] F. A. GRAYBILL AND A. W. WORTHAM, "A note on uniformly best unbiased estimators for variance components", *J. Amer. Stat. Assn.*, Vol. 51 (1956), pp. 266–268.

[4] L. H. HERBACH, "Topics in analysis of variance: A. Optimum properties of tests for model II, B. Generalizations of model II" (Abstract), *Ann. Math. Stat.* Vol. 24 (1953), p. 137.

[5] L. H. HERBACH, "Properties of model II—Type analysis of variance tests, A: Optimum nature of the F-test for model II in the Balanced Case", *Ann. Math. Stat.* Vol. 30 (1959), pp. 939–959.

[6] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation, Part I" *Sankhya*, Vol. 10 (1950), pp. 305–340.

[7] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation, Part II", *Sankhya*, Vol. 15 (1955), pp. 219–236.

[8] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, 1941.

# THE MOST-ECONOMICAL CHARACTER OF SOME BECHHOFER AND SOBEL DECISION RULES[1]

BY WM. JACKSON HALL

*University of North Carolina*

**1. Introduction.** R. E. Bechhofer [1] has considered a single-sample multiple-decision procedure for choosing, among a group of normal populations with common known variances, that population with the largest mean, and, with M. Sobel [2], a procedure for choosing the normal population with the smallest variance. Several other analogous problems have also been considered.[2] They suggest, with only intuitive justification, choosing the population with the largest (smallest) sample mean (variance), and give tables for finding the minimum sample size (assumed equal for all populations) which will guarantee a correct decision with prescribed probability when the extreme population parameter is sufficiently distinct from the others. This paper gives justification for a wide class of such procedures, proving that no other rules can meet this guarantee with a smaller (fixed) sample size; that is, such rules are *most economical* [4].

Proof of the most-economical character of these rules is achieved by proving their minimax character when a suitable loss function is introduced. R. R. Bahadur and L. A. Goodman [5] have considered a class of multiple-decision rules which they have called *impartial* (invariant under permutations of the populations). Their results are applicable to such problems of choosing the best population and imply that Bechhofer and Sobel's rules are minimax rules (in fact, uniformly minimum risk rules) among the class of impartial decision rules. The present paper removes this restriction of impartiality. Thus, in the present context, impartiality is no restriction when looking for minimax rules, as is well-known to be the case for certain other kinds of invariance.

The main result is stated in Section 2 and proved in Section 3. It is applicable to any analogous problem of choosing the population with the most extreme parameter when, for each sample, there is a numerical sufficient statistic with a *monotone likelihood ratio*[3] and the (numerical) parameter is a *location* or *scale* (but not range) *parameter*[3] in the distribution of the statistic. The theorem is applicable to Bechhofer's procedure and the corollary to Bechhofer and Sobel's. (In the latter example, if the means are unknown, it will be necessary to invoke invariance under changes in scale.) In Section 4, the result is further extended to problems

Received November 23, 1958; revised March 26, 1959.

[1] This research was supported by the United States Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] For a review, see the introduction in [3].

[3] For definition, see [6], for example.

of ranking the populations according to the parameter values, or of grouping them by ranks, as formulated by Bechhofer [1].

The requirement that the parameter be one of location or scale is dropped in Section 5. Then the *guarantee* holds only at a *specified location*; for many problems, a *least favorable location* can be determined so that the guarantee can be made to hold irrespective of location. For example, the procedures of M. Sobel and M. J. Huyett [7] for choosing the largest of several binomial parameters are proved to be most economical. In Section 6, the broader optimality of these latter procedures is discussed.

These results, some of which appeared in [8], are obtained from application of *most economical decision theory* [4]. As indicated by Bechhofer [1], if the populations differ in a known way (normal populations with different known variances, for example), optimal allocation of the sample sizes is apparently exceedingly complex; such problems are not treated here.

**2. Theorem.**

(i) *Let* $\{f_\theta\}$, $\theta \varepsilon \Omega \subset R_1$, *be a homogeneous class of density functions*[4] *w.r.t. a fixed measure. Let* $\{X_{ij}\}$ $(i = 1, \cdots, m; j = 1, \cdots, n)$ *denote mn independent random variables where* $X_{ij}$ *has the density function* $f_{\theta_i}$, $\theta_i \varepsilon \Omega$, $i = 1, \cdots, m$, *and let* $\theta_{[1]} \leqq \cdots \leqq \theta_{[m]}$ *be the ordered values of the* $\theta_i$*'s.* Set $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)$.

(ii) *Suppose* $t_i = t_i(x_{i1}, \cdots, x_{in})$ *is a numerical sufficient statistic for* $(X_{i1}, \cdots, X_{in})$, *that* $t_i$ *has a monotone likelihood ratio, and that* $\theta_i$ *is a location parameter in the induced distribution of* $t_i$ $(i = 1, \cdots, m)$.

(iii) *Let* $D_n$ *denote any decision rule for choosing which* $\theta_i$ *is* $\theta_{[m]}$ *based on an observation on the mn random variables* $\{X_{ij}\}$, *and let* $D_n^0$ *denote that* $D_n$ *which chooses as* $\theta_{[m]}$ *that* $\theta_i$ *corresponding to the largest of the* $t_i$*'s with ties broken by randomization. Suppose N is the least n for which*

$$\begin{aligned}(1)\qquad &\int F_n^{m-1}(t + \delta - 0)\, dF_n(t) + \sum_{r=2}^{m} \frac{1}{r}\binom{m-1}{r-1} \\ &\cdot \int [F_n(t+\delta) - F_n(t + \delta - 0)]^{r-1}[F_n(t+\delta-0)]^{m-r}\, dF_n(t) \geqq \gamma \\ &\qquad\qquad (\delta > 0, 0 < \gamma < 1)\end{aligned}$$

*where* $F_{\theta,n}(t) = F_n(t - \theta)$ *is the c.d.f. of t with parameters* $\theta$ *and n.*

*Then* $D_N^0$ *satisfies*

(a) $Pr\{$*correct decision using* $D_n \mid \boldsymbol{\theta}\} \geqq \gamma$ *for all* $\boldsymbol{\theta}$ *for which* $\theta_{[m]} - \theta_{[m-1]} \geqq \delta$ *and*

(b) *there does not exist a decision rule* $D_n$ *satisfying* (a) *with* $n < N$.

Corollary: *Replace in* (i) "$R_1$" *by* "*positive* $R_1$"; *replace in* (ii) "*location*" *by* "*scale*"; *replace in* (iii) "$t + \delta$" *by* "$t\delta$", "$\delta > 0$" *by* "$\delta > 1$", "$F_n(t - \theta)$" *by* "$F_n(t/\theta)$"; *replace in* (a) "$\theta_{[m]} - \theta_{[m-1]}$" *by* "$\theta_{[m]}/\theta_{[m-1]}$".

*Note*: The summation term in (1) accommodates the possibility of ties—when $r = 2, \cdots, m$ $t$-values may be largest—and drops out if $F_{\theta,n}$ is absolutely con-

[4] The region of positive density is independent of $\theta$.

tinuous; hereafter, for simplicity of presentation, we make this assumption and thereby replace (1) by

$$(1') \qquad \int F_n^{m-1}(t+\delta)\, dF_n(t) \geqq \gamma.$$

**3. Proof of theorem.** Set $\omega_i = \{\boldsymbol{\theta} \mid \theta_i = \theta_{[m]}, \theta_{[m]} - \theta_{[m-1]} \geqq \delta\}$, and $p_i(\boldsymbol{\theta}) = \Pr\{\text{choosing } \theta_i \text{ using } D_n \mid \boldsymbol{\theta}\}$, $i = 1, \cdots, m$. Then (a) is equivalent to: $p_i(\boldsymbol{\theta}) \geqq \gamma$ for $\boldsymbol{\theta} \,\varepsilon\, \omega_i$ $(i = 1, \cdots, m)$.

Let $\lambda_i$ be a distribution over $\omega_i$ which assigns probability one to the $\boldsymbol{\theta}$-point with all coordinates equal to $\theta_0$ (arbitrary) except the $i$th coordinate which equals $\theta_0 + \delta$; i.e., $\theta_{[1]} = \theta_{[m-1]} = \theta_0 = \theta_i - \delta$. Denote this point $\boldsymbol{\theta}_i$.

We first show that, for $n$ fixed, $D_n^0$ is minimax for choosing among $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_m$ where the loss function is $-1/\gamma$ if a correct decision is made and zero otherwise, and that, when using $D_n^0$, $p_i(\boldsymbol{\theta}_i) = \int F_n^{m-1}(t+\delta)\, dF_n(t)$ for all $i$. Secondly, we show that the $\lambda_i$'s are least favorable in the sense that $\inf_{\omega_i} p_i(\boldsymbol{\theta}) = p_i(\boldsymbol{\theta}_i) = \int_{\omega_i} p_i(\boldsymbol{\theta})\, d\lambda_i$, as shown in the special case of Bechhofer in [1]. Application of Theorems 7 and 9 from [4] completes the proof of the theorem. The corollary may be proved by applying a log transform to $t$, $\theta$, and $\delta$.

1.° According to well-known results of Wald (e.g., see Section 1.B of [4]), a minimax rule for choosing among the $\boldsymbol{\theta}_i$'s with the specified loss is one which chooses $\theta_i$ as the largest $\theta$ if $a_i h_i \geqq a_j h_j$ for all $j$ where $h(\mathbf{t}, \boldsymbol{\theta})$ is the joint density of $t_1, \cdots, t_m$ when the parameter is $\boldsymbol{\theta}$, $h_j = h(\mathbf{t}, \boldsymbol{\theta}_j)$, and $a_1, \cdots, a_m$ are positive constants chosen so that $p_1(\boldsymbol{\theta}_1) = \cdots = p_m(\boldsymbol{\theta}_m)$. Denoting the density of $F$ by $g$ and of $F_\theta$ by $g_\theta$ (dropping the subscript $n$ assumed fixed), $h(\mathbf{t}, \boldsymbol{\theta}) = g_{\theta_1}(t_1) g_{\theta_2}(t_2) \cdots g_{\theta_m}(t_m)$ so that $a_i h_i \geqq a_j h_j$ implies

$$(2) \qquad a_i g_{\theta_0+\delta}(t_i) g_{\theta_0}(t_j) \geqq a_j g_{\theta_0}(t_i) g_{\theta_0+\delta}(t_j),$$

or equivalently, since $\theta$ is a location parameter, the subscripts on the $g$'s can be subtracted from the arguments. Denoting $r(t) = g_{\theta_0+\delta}(t)/g_{\theta_0}(t)$ for fixed $\theta_0$ and $\delta$, defined throughout the region of positive density for $t$, (2) implies $r(t_j) \leqq r(t_i)a_i/a_j$. Since $t$ has a monotone likelihood ratio, $r(t)$ increases with $t$, the inverse function exists, and (2) may be written $t_j \leqq r^{-1}[r(t_i)a_i/a_j]$. Therefore, the probability that the minimax rule chooses $\theta_i$ as largest when $\boldsymbol{\theta} = \boldsymbol{\theta}_i$ is

$$(3) \qquad \begin{aligned} p_i(\boldsymbol{\theta}_i) &= \Pr\{a_i h_i \geqq a_j h_j \quad \text{for all} \quad j \mid \boldsymbol{\theta} = \boldsymbol{\theta}_i\} \\ &= \mathcal{E} \Pr\{a_i h_i \geqq a_j h_j \quad \text{for all} \quad j \mid t_i = y, \boldsymbol{\theta} = \boldsymbol{\theta}_i\} \\ &= \int \prod_{j \neq i} F_{\theta_0, n}\{r^{-1}[r(y)a_i/a_j]\}\, dF_{\theta_0+\delta, n}(y). \end{aligned}$$

This is independent of $i$ if $a_1 = a_2 = \cdots = a_m$, in which case $\theta_i$ is chosen if $h_i$ is largest. Because of the monotone property of $h_i$, the minimax rule is thus $D_n^0$. Upon setting the $a_i$'s equal and transforming $t = y - \theta_0 - \delta$, (3) becomes $p_i(\boldsymbol{\theta}_i) = \int F_n^{m-1}(t+\delta)\, dF_n(t)$, and is thus independent of the choice of $\theta_0$.

2.° Similarly to (3) above, for $D_n^0$ we have $p_i(\boldsymbol{\theta}) = \Pr\{t_i \geqq t_j \text{ for all } j \mid \boldsymbol{\theta}\} =$

$\int \Pi_{j \neq i} F_{\theta_j, n}(y)\, dF_{\theta_i, n}(y) = \int \Pi_{j \neq i} F_n(u + \theta_i - \theta_j)\, dF_n(u)$ which increases with $\theta_i - \theta_j$ for each $j \neq i$. For $\boldsymbol{\theta} \varepsilon \omega_i$, $\theta_i - \theta_j \geqq \delta$ so that the infimum over $\omega_i$ of $p_i$ is attained at $\theta_i - \theta_j = \delta$ for $j \neq i$ and, in particular, at $\boldsymbol{\theta} = \boldsymbol{\theta}_i$.

**4. Extension to procedures for grouping by ranks.** The results in Section 2 can be extended to the problem of ranking the populations according to their $\theta$-values, or more generally, of selecting the $m_s$ "best" populations, the $m_{s-1}$ "second best", etc., the $m_1$ "worst" populations, given $m_1, \cdots, m_s (s \leqq m, \sum m_i = m)$—the "general goal" expressed by Bechhofer in Section 3.B of [1]. The rule $D_N^0$ is to rank according to $t$-values, choose $N$ by a rule analogous to (1) or (1′) (see [1]), and then the probability of a correct grouping will be at least $\gamma$ when the groups are sufficiently far apart. Most economical theory is used for discriminating among the $m!/(m_1!\, m_2! \cdots m_s!)$ possible alternative decisions. The proof differs little except for the notational complexities.

**5. Extension to other ordering parameters.** Intuitively, a procedure which ranks the $\theta$'s according to the values of the sufficient statistic $t$ should be optimal whenever $\theta$ is some kind of *ordering parameter* in the distribution of $t$. That $t$ should have a monotone likelihood ratio is such an ordering requirement. A less stringent requirement is that the c.d.f. of $t$ be monotone in $\theta$ for all $t$; that is, denoting by $T_i$ a random variable with distribution parameter $\theta_i (\theta_1 < \theta_2)$, $\Pr\{T_2 > t\} \geqq \Pr\{T_1 > t\}$ for all $t$, in which case $T_2$ is said to be *stochastically larger* than $T_1$. E. L. Lehmann has shown (Theorem 1 in [9]) that a monotone likelihood ratio assumption implies the latter type of ordering. That $\theta$ be a location parameter is an additional ordering requirement—that $F_\theta$ be a particular kind of monotone function, namely $F(t - \theta)$. It was required in the theorem so that the probability in (a) could be computed on the condition that the best population was sufficiently *distant* from the second best without regard to the *location* of the best population; that $F_\theta$ be monotone was also required, but this follows from the monotone likelihood ratio assumption. Thus, the location requirement in (ii) can be removed by adding it in (a), so that replacing $F(t + \delta)$ by $F_{\theta_0 - \delta}(t)$ and $dF$ by $dF_{\theta_0}$ in (1) and (1′) and replacing (a) by (a′) in which the inequality is required to hold for all $\boldsymbol{\theta}$ for which $\theta_{[m]} = \theta_0$ and $\theta_{[m]} - \theta_{[m-1]} \geqq \delta$ for some specified value $\theta_0$ of the parameter, the theorem and proof remain valid. The guarantee of a correct decision is only calculated at one location, specified by $\theta_0$.

In many such problems, it will be possible to find a least favorable location $\theta_0$, i.e., a value of $\theta$ which minimizes $\int F_{\theta-\delta}^{m-1}(t)\, dF_\theta(t)$, in which case (a) need not be replaced by (a′). Sufficient conditions are that $\Omega$ be bounded and closed. If not, it may be possible to find a *least favorable sequence*, applying Theorem 8 of [4].

In this revised form, the theorem applies to all such problems of choosing the best population whenever there is a numerical sufficient statistic with a monotone likelihood ratio, and therefore, in particular, if its distribution is in the exponential family. Thus, it holds for Sobel and Huyett's procedures [7],

the condition (a) corresponding to their "original specification" using a least favorable $\theta_0$, and (a′) to their "alternative specification."

**6. A distribution-free extension.** It can be shown that Sobel and Huyett's procedure is optimal not only for choosing the best binomial population but for the more general problem they describe of choosing the population with the largest "survival probability", with no parametric specification of the underlying distributions. If the distributions differ only in location, then the problem is equivalent to that of choosing the population with the largest median. This application is adapted from an example by W. Hoeffding [10].

Let the class of density functions under consideration include all densities, $f$, w.r.t. a fixed measure $\mu$ on the real line such that $0 < \mu(\{x \geqq a\}) < 1$ for some specified $a$. The $\{X_{ij}\}$ are assumed independent with $\theta_i \equiv \Pr\{X_{ij} \geqq a\}$, constant over $j = 1, \cdots, n (i = 1, \cdots, m)$. If $X_{ij}$ represents a lifetime, then $\theta_i$ is the probability of survival to age $a$, and none of the distributions need coincide except in their $\theta$-values.

Extension of the theorem can be accomplished as indicated briefly here: Subsets $\{\omega_i\}$ of density functions are specified in terms of the $\theta$'s as in Section 3; *a priori* distributions over these sets are specified, somewhat as in example 5 in [4], which reduces the problem to that of choosing the best of $m$ binomial distributions. The decision procedure is to choose $\theta_k$ as the largest of the $\theta_i$'s if more of the $x_{kj}$'s exceed $a$ than do the $x_{ij}$'s for any other $i$. That these *a priori* distributions are least favorable follows as in Section 3, using Theorem 7 from [4], and noting that the probability of a correct decision depends only on the $\theta$-values. A least favorable $\theta_0$ can be chosen, if desired, as in [7]. Thus, the procedure of Sobel and Huyett is most economical for this distribution-free problem in the sense that (a), or (a′), and (b) are satisfied.

## REFERENCES

[1] Robert E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–39.

[2] Robert E. Bechhofer and Milton Sobel, "A single-sample multiple decision procedure for ranking variances of normal populations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 273–289.

[3] Robert E. Bechhofer, "A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs," *Biometrics*, Vol. 14 (1958), pp. 408–429.

[4] Wm. Jackson Hall, "Most economical multiple-decision rules," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1079–1094.

[5] Raghu Raj Bahadur and Leo A. Goodman, "Impartial decision rules and sufficient statistics," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 553–562.

[6] David Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[7] Milton Sobel and Marilyn J. Huyett, "Selecting the best one of several binomial populations," *Bell System Tech. Jour.*, Vol. 36 (1957), pp. 537–576.

[8] Wm. Jackson Hall, "An optimum property of Bechhofer's single-sample multiple-decision procedure for ranking means and some extensions," *Institute of Statistics Mimeograph Series*, No. 118, University of North Carolina, September 1954.

[9] E. L. Lehmann, "Ordered families of distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 399–419.

[10] Wassily Hoeffding, " 'Optimum' nonparametric tests," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 83–92.

# THE ADMISSIBILITY OF PITMAN'S ESTIMATOR OF A SINGLE LOCATION PARAMETER[1]

By Charles Stein

*Stanford University and University of California, Berkeley*[2]

**1. Introduction.** Pitman [1] gave a thorough discussion of the problem of estimating the location and scale parameters of a distribution which is known except for one or both of these parameters. In particular, if $X_1 \cdots X_n$ are real random variables independently and identically distributed according to the density $r(x - \xi)$ (with respect to Lebesgue measure), where $\xi$ is unknown but the function $r$ is known, Pitman shows that the estimator

$$\hat{\xi}_0(X_1, \cdots, X_n) = \frac{\int \xi \prod r(X_i - \xi)\, d\xi}{\int \prod r(X_i - \xi)\, d\xi} \tag{1.1}$$

is the best translation-invariant estimator in the sense that it minimizes $\mathcal{E}_\xi[\hat{\xi}(X_1 \cdots X_n) - \xi]^2$ among all estimators $\hat{\xi}$ for which

$$\hat{\xi}(x_1 + c, \cdots, x_n + c) = \hat{\xi}(x_1, \cdots, x_n) + c \tag{1.2}$$

for all $x_1, \cdots, x_n$ and $c$. Girshick and Savage [2] showed that $\hat{\xi}_0$ is minimax in the class of all estimators (not restricted by (1.2)) and this also follows from the later more general results of Kudo [3] and Kiefer [4]. Karlin [5] has shown that under certain conditions $\hat{\xi}_0$ is admissible, that is, if $\hat{\xi}$ is any estimator for which

$$E_\xi(\hat{\xi}(X_1, \cdots, X_n) - \xi)^2 \leqq E_\xi(\hat{\xi}_0(X_1, \cdots, X_n) - \xi)^2 \tag{1.3}$$

for all $\xi$, then equality holds for all $\xi$. Since his conditions are fairly strong, and his method somewhat special, it seems desirable to present an alternative proof. Theorem 1 of Section 2, when reformulated for the present slightly special case, becomes

Theorem. *If*

$$\int \prod r(x_i) \left\{ \frac{\int \xi^2 \prod r(x_i - \xi)\, d\xi}{\int \prod r(x_i - \xi)\, d\xi} - \left( \frac{\int \xi \prod r(x_i - \xi)\, d\xi}{\int \prod r(x_i - \xi)\, d\xi} \right)^2 \right\}^{3/2} \prod dx_i < \infty \tag{1.4}$$

*then* $\hat{\xi}_0$ *defined by* (1.1) *is admissible.*

Received August 25, 1958.

[1] This work was sponsored by the Office of Naval Research.

[2] The author is now at Stanford University.

The condition (1.4) is not very strong. For example, if there is any translation-invariant estimator $\hat{\xi}$ for which $\mathcal{E}_\xi | \hat{\xi} - \xi |^3 < \infty$, then (1.4) holds. For the Cauchy distribution $r(x) = 1/\pi (1 + x^2)$ with $n \geqq 7$, this is true with $\hat{\xi}$ equal to the sample median.

The proof is given by a method first used by Blyth [6], and the result seems to be the best possible obtainable by this method. Here, as in Lehmann and Stein [7], roughly speaking, the theorem requires one more moment than is clearly relevant. In [7] a first moment is required, although it is a testing problem, and here, a third moment rather than a second. It would be interesting to know whether conditions of this type are necessary. Essentially the same method will be applied in a paper, now being prepared, to the problem of estimating two unknown location parameters with quadratic loss. There it is necessary to vary the form, as well as the scale, of the *a priori* distribution (see the argument around (2.16)). The bivariate normal case has already been treated by the author in [8]. For three or more translation parameters with positive definite quadratic loss, Pitman's estimator is not admissible. This was proved in the normal case in [8]. While it is of some theoretical interest to prove the admissibility of the natural estimator when it is admissible, the careful study of other estimators when the natural estimator is not admissible has greater practical value.

It may be useful to indicate the correspondence between the notation used in this introduction and that of the slightly more general problem treated in the remainder of the paper. Let $\mathcal{Y}$ be the $n - 1$ dimensional real coordinate space, $\mathcal{C}$ the $\sigma$-algebra of all Borel subsets of $\mathcal{Y}$ and $\nu$ the distribution of $Y$ defined by (1.9).

$$f(y) = \frac{\int xr(x)r(x + y_1) \cdots r(x + y_{n-1})\, dx}{\int r(x)r(x + y_1) \cdots r(x + y_{n-1})\, dx} \tag{1.5}$$

and

$$g(y) = \int r(x)r(x + y_1) \cdots r(x + y_{n-1})\, dx, \tag{1.6}$$

where

$$y = (y_1, \cdots, y_{n-1}). \tag{1.7}$$

Also, let

$$p(x, y) = \frac{r(x + f(y))r(x + f(y) - y_1) \cdots r(x + f(y) - y_{n-1})}{g(y)}. \tag{1.8}$$

Then conditions (2.1) and (2.2) are satisfied by $p$, and (1.4) reduces to (2.3). If we define the random point $(X, Y)$ by

$$\begin{aligned} Y_1 &= X_2 - X_1, \\ &\vdots \\ Y_{n-1} &= X_n - X_1, \end{aligned} \tag{1.9}$$

$$X = X_1 - f(Y_1, \cdots, Y_{n-1}), \tag{1.10}$$

then the estimate $X$, proved admissible in Section 2, is seen to reduce to $\xi_0(X_1, \cdots, X_n)$.

**2. The results.** Let $\mathfrak{B}$ be the $\sigma$-algebra of all Borel subsets of the real line $\mathfrak{X}$, and $\mathfrak{C}$ a $\sigma$-algebra of subsets of a set $\mathfrak{Y}$. Let $\mu$ be Lebesgue measure on $\mathfrak{B}$ and $\nu$ a probability measure on $\mathfrak{C}$. Let $p$ be a nonnegative valued $\mathfrak{BC}$ measurable function on $\mathfrak{X} \times \mathfrak{Y}$ such that

$$\left.\begin{aligned} &\int p(x, y)\, dx = 1 &&(2.1)\\ &\int xp(x, y)\, dx = 0 &&(2.2)\end{aligned}\right\} \text{ for all } y$$

$$\int d\nu(y) \left(\int x^2 p(x, y)\, dx\right)^{3/2} < \infty, \tag{2.3}$$

where we write $dx$ instead of $d\mu(x)$. Then of course $p$ is a probability density with respect to $\mu\nu$. We shall prove

THEOREM 1. *Under the above hypotheses, if we observe $(X, Y)$ distributed so that, for some unknown $\xi$, $(X - \xi, Y)$ has probability density $p$ with respect to $\mu\nu$, then $X$ is an admissible estimator of $\xi$ with squared error as loss.*

In other words, if $\varphi$ is any $\mathfrak{BC}$ measurable function on $\mathfrak{X} \times \mathfrak{Y}$ such that

$$\begin{aligned}&\int d\nu(y) \int [\varphi(x, y) - \xi]^2 p(x - \xi, y)\, dx \\ &\quad \leq \int d\nu(y) \int (x - \xi)^2 p(x - \xi, y)\, dx = \int d\nu(y) \int x^2 p(x, y)\, dx\end{aligned} \tag{2.4}$$

for all $\xi$, then the two sides of (2.4) are identically equal. Actually we prove the trivially stronger result that $\varphi(x, y) = x$ almost everywhere $(\mu\nu)$. One might hope to prove this result under the condition

$$\int d\nu(y) \int x^2 p(x, y)\, dx < \infty, \tag{2.3$'$}$$

which is weaker than (2.3). Of course (2.3$'$) is necessary, for otherwise we could take $\varphi(x, y) \equiv 0$.

We shall derive Theorem 1 from a slightly more general but weaker theorem. With $\mathfrak{X}$, $\mathfrak{Y}$, $\mathfrak{B}$, $\mathfrak{C}$, $\mu$, $\nu$ as before, let $P$ be a nonnegative valued $\mathfrak{BC}$ measurable function on $\mathfrak{X} \times \mathfrak{Y}$ such that, for each $y$, $P(\cdot, y)$ is a cumulative distribution function and

$$\int x\, d_x P(x, y) = 0 \tag{2.5}$$

$$\int d\nu(y) \left(\int x^2\, d_x P(x, y)\right)^{3/2} < \infty \tag{2.6}$$

THEOREM 2. *Under the above hypotheses, if we observe* $(X, Y)$ *distributed so that, for some unknown* $\xi$, $Y$ *is distributed according to* $\nu$ *and the conditional cumulative distribution function of* $X - \xi$ *given* $Y$ *is* $P(\cdot, Y)$, *then* $X$ *is an almost admissible estimator of* $\xi$ *with squared error as loss. That is, if* $\varphi$ *is any* $\mathcal{BC}$ *measurable function on* $\mathfrak{X} \times \mathcal{Y}$ *such that*

$$\begin{aligned}(2.7)\qquad &\int d\nu(y) \int [\varphi(x, y) - \xi]^2\, d_x P(x - \xi, y) \\ &\leqq \int d\nu(y) \int (x - \xi)^2\, d_x P(x - \xi, y) = \int d\nu(y) \int x^2\, d_x P(x, y)\end{aligned}$$

*for all* $\xi$, *then the two sides are equal for almost all* $\xi$.

By a familiar argument, we observe that Theorem 1 follows from Theorem 2, if we put

$$(2.8)\qquad P(x, y) = \int_{-\infty}^{x} p(t, y)\, dt.$$

If $\varphi$ satisfies the hypotheses of Theorem 1, it also satisfies those of Theorem 2 and we conclude that in (2.4) equality holds for almost all $\xi$. Now suppose that contrary to the conclusion of Theorem 1,

$$(2.9)\qquad \mu\nu(S) > 0,$$

where

$$(2.10)\qquad S = \{(x, y) : \varphi(x, y) \neq x\}.$$

Then, for all $\xi$ in a set $T$ of positive measure,

$$(2.11)\qquad \int d\nu(y) \int_{S_y} p(x - \xi, y)\, dx > 0,$$

where $S_y = \{x : \varphi(x, y) \neq x\}$, since

$$\begin{aligned}(2.12)\qquad &\int d\xi \int d\nu(y) \int_{S_y} p(x - \xi, y)\, dx \\ &= \int d\nu(y) \int_{S_y} dx \int p(x - \xi, y)\, d\xi = \int d\nu(y) \int_{S_y} dx = \mu\nu(S).\end{aligned}$$

Let

$$(2.13)\qquad \varphi_0(x, y) = \tfrac{1}{2}(x + \varphi(x, y)).$$

Then

$$(2.14)\qquad [\varphi_0(x, y) - \xi]^2 \leqq \tfrac{1}{2}\{[\varphi(x, y) - \xi]^2 + (x - \xi)^2\}$$

with strict inequality whenever $\varphi(x, y) \neq x$. It follows that we have strict inequality in (2.4) and thus in (2.7) for all $\xi \in T$ contradicting the conclusion of Theorem 2. An example given by Blackwell [9] with $\mathcal{Y}$ reducing to a point

and $P$ concentrated on a finite set shows that in Theorem 2 we cannot conclude admissibility.

To prove Theorem 2 we suppose the conclusion does not hold, that is, we suppose (2.7) holds with strict inequality for $\xi$ in a set $S$ having positive Lebesgue measure. For $\epsilon > 0$, let $S_\epsilon$ be the set of $\xi$ for which

$$\int d\nu(y) \int [\varphi(x, y) - \xi]^2 \, dP(x - \xi, y) \leqq \int d\nu(y) \int x^2 \, dP(x, y) - \epsilon. \tag{2.15}$$

Since $S = \cup S_\epsilon$, $S_\epsilon$ will have positive Lebesgue measure for sufficiently small $\epsilon$, and we suppose $\epsilon$ chosen so that $\mu(S_\epsilon) > 0$. Since $S_\epsilon$ (like any measurable set) is of density 1 at almost all points of itself (see for example Titchmarsh [10], p. 371), there exists $\kappa > 0$ and an interval $I = (a - \kappa, a + \kappa)$ such that the set of $\xi \varepsilon I$ for which (2.15) holds has Lebesgue measure $\geqq \kappa$. There is no real loss of generality in assuming $I = (-\kappa, \kappa)$. Now we assign to $\xi$ an *a priori* density $(1/\sigma)q(\xi/\sigma)$, taking for simplicity of computation

$$q(\xi) = \frac{1}{\pi(1 + \xi^2)}. \tag{2.16}$$

From (2.7), and the fact that (2.15) holds for a set of measure $\geqq \kappa$ in $(-\kappa, \kappa)$, it follows that

$$\mathcal{E}[\varphi(X, Y) - \xi]^2 \leqq \int d\nu(y) \int x^2 \, dP(x, y) - \frac{\epsilon\kappa}{2\pi\sigma} \tag{2.17}$$

for sufficiently large $\sigma$, where $\xi$ has the indicated *a priori* distribution, and the conditional distribution of $(X, Y)$ given $\xi$ is that indicated before Theorem 2. However, we shall show that under the same distribution

$$\inf_{\psi} E[\psi(X, Y) - \xi]^2 \geqq \int d\nu(y) \int x^2 \, dP(x, y) - \frac{f(\sigma)}{\sigma}, \tag{2.18}$$

where

$$\lim_{\sigma \to \infty} f(\sigma) = 0. \tag{2.19}$$

For sufficiently large $\sigma$ this contradicts (2.17).

We shall find the formula

$$\begin{aligned} &\int d\nu(y) \int x^2 \, dP(x, y) - \inf_{\psi} E[\psi(X, Y) - \xi]^2 \\ &\qquad = \frac{1}{\sigma} \int d\nu(y) \int dx \frac{\left[ \int \eta q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y) \right]^2}{\int q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y)} \end{aligned} \tag{2.20}$$

useful in proving (2.18). To prove (2.20) we first observe that

$$\inf_{\psi} E[\psi(X, Y) - \xi]^2 = \inf_{\psi} EE\{[\psi(X, Y) - \xi]^2 \mid X, Y\} \tag{2.21}$$
$$= E\{E[\xi \mid (X, Y)] - \xi\}^2,$$

so that

$$\int d\nu(y) \int x^2 \, dP(x, y) - \inf E[\psi(X, Y) - \xi]^2$$
$$= E(X - \xi)^2 - E\{E[\xi \mid (X, Y)] - \xi\}^2$$
$$= E\{X^2 - 2\xi X + (E[\xi \mid (X, Y)])^2\} = E\{X - E[\xi \mid (X, Y)]\}^2$$
$$= \frac{1}{\sigma} \int q\left(\frac{\xi}{\sigma}\right) d\xi \int d\nu(y) \int$$
$$\cdot \left\{ x - \frac{\int \xi' q\left(\frac{\xi'}{\sigma}\right) d_{\xi'} P(x - \xi', y)}{\int q\left(\frac{\xi'}{\sigma}\right) d_{\xi'} P(x - \xi', y)} \right\}^2 d_x P(x - \xi, y)$$
$$= \frac{1}{\sigma} \int d\nu(y) \int q\left(\frac{\xi}{\sigma}\right) d\xi \int \left[ \frac{\int \eta q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y)}{\int q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y)} \right]^2 d_x P(x - \xi, y)$$
$$= \frac{1}{\sigma} d\nu(y) \int q\left(\frac{\xi}{\sigma}\right) d\xi \int \left[ \frac{\int \eta q\left(\frac{\xi + \eta' - \eta}{\sigma}\right) dP(\eta, y)}{\int q\left(\frac{\xi + \eta' - \eta}{\sigma}\right) dP(\eta, y)} \right]^2 dP(\eta', y) \tag{2.22}$$
$$= \frac{1}{\sigma} \int d\nu(y) \int dP(\eta', y) \int \left[ \frac{\int \eta q\left(\frac{\xi + \eta' - \eta}{\sigma}\right) dP(\eta, y)}{\int q\left(\frac{\xi + \eta' - \eta}{\sigma}\right) dP(\eta, y)} \right]^2 q\left(\frac{\xi}{\sigma}\right) d\xi$$
$$= \frac{1}{\sigma} \int d\nu(y) \int dP(\eta', y) \int \left[ \frac{\int \eta q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y)}{\int q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y)} \right]^2 q\left(\frac{x - \eta'}{\sigma}\right) dx$$
$$= \frac{1}{\sigma} \int d\nu(y) \int dx \frac{\left[ \int \eta q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y) \right]^2}{\int q\left(\frac{x - \eta}{\sigma}\right) dP(\eta, y)}.$$

Comparing this result with (2.17) and (2.18), we see that, in order to complete the proof of Theorem 2, we need only show that

$$\lim_{\sigma\to\infty} \int d\nu(y) \int \frac{\left[\int \eta q\left(\frac{x-\eta}{\sigma}\right) dP(\eta, y)\right]^2}{\int q\left(\frac{x-\eta}{\sigma}\right) dP(\eta, y)} dx = 0. \tag{2.23}$$

In order to prove (2.23) we consider the integral

$$\Phi(P, \sigma) = \int dx \frac{\left[\int \eta q\left(\frac{x-\eta}{\sigma}\right) dP(\eta)\right]^2}{\int q\left(\frac{x-\eta}{\sigma}\right) dP(\eta)} \tag{2.24}$$

and the function

$$\psi(\lambda, \sigma) = \sup_{P \in \mathfrak{U}_\lambda} \Phi(P, \sigma), \tag{2.25}$$

where $\mathfrak{U}_\lambda$ is the set of probability measures $P$ for which

$$\int \eta \, dP(\eta) = 0 \tag{2.26}$$

$$\int \eta^2 \, dP(\eta) = \lambda. \tag{2.27}$$

As indicated earlier, we take

$$q(\xi) = \frac{1}{\pi(1+\xi^2)}, \tag{2.28}$$

but the basic formulas hold for an arbitrary $q$. We first observe that

$$\psi(\lambda, \sigma) = \sigma^3 \psi\left(\frac{\lambda}{\sigma^2}, 1\right), \tag{2.29}$$

since

$$\begin{aligned} \Phi(P, \sigma) &= \sigma^3 \int d\left(\frac{x}{\sigma}\right) \frac{\left[\int \frac{\eta}{\sigma} q\left(\frac{x-\eta}{\sigma}\right) dP(\eta)\right]^2}{\int q\left(\frac{x-\eta}{\sigma}\right) dP(\eta)} \\ &= \sigma^3 \int dx \frac{\left[\int \eta q(x-\eta) \, dP(\eta\sigma)\right]^2}{\int q(x-\eta) \, dP(\eta\sigma)}. \end{aligned} \tag{2.30}$$

We observe also that

$$\psi(\lambda, 1) \leqq \lambda, \tag{2.31}$$

a bound which will be useful only for large $\lambda$. This follows from the convexity of $\Phi$, or from

$$\Phi(P, 1) = \int x^2\, dP(x) - \inf_{\psi} \mathcal{E}[\psi(X) - \xi]^2, \tag{2.32}$$

with $\xi$ distributed according to $q$, and $X - \xi$ given $\xi$ according to $P$, which is essentially (2.22).

For $\lambda \leqq \frac{1}{2}$,

$$\begin{aligned}\int q(x - \eta)\, dP(\eta) &\geqq P[-1, 1] \inf_{z \epsilon [-1,1]} q(x - z) \\ &\geqq \frac{3}{4} \cdot \frac{2}{5} \frac{1}{\pi(1 + x^2)} = \frac{3}{10\pi(1 + x^2)}\end{aligned} \tag{2.33}$$

by Chebyshev's inequality. Also

$$\begin{aligned}\left[\int \eta q(x - \eta)\, dP(\eta)\right]^2 &= \left\{\int \eta[q(x - \eta) - q(x)]\, dP(\eta)\right\}^2 \\ &\leqq \int \eta^2\, dP(\eta) \int [q(x - \eta) - q(x)]^2\, dP(\eta)\end{aligned} \tag{2.34}$$

by Schwarz's inequality. Thus

$$\begin{aligned}\Phi(P, 1) &= \int dx \frac{\left[\int \eta q(x - \eta)\, dP(\eta)\right]^2}{\int q(x - \eta)\, dP(\eta)} \\ &\geqq \frac{10}{3} \int dx(1 + x^2) \left(\int \eta^2\, dP(\eta)\right) \int \left[\frac{1}{1 + (x - \eta)^2} - \frac{1}{1 + x^2}\right]^2 dP(\eta) \\ &= \frac{10}{3}\left(\int \eta^2\, dP(\eta)\right) \int dP(\eta) \int \left[\frac{1}{1 + (x - \eta)^2} - \frac{1}{1 + x^2}\right]^2 (1 + x^2)\, dx.\end{aligned} \tag{2.35}$$

But

$$\begin{aligned}&\int \left[\frac{1}{1 + (x - \eta)^2} - \frac{1}{1 + x^2}\right]^2 (1 + x^2)\, dx \\ &\quad = \int \left[\frac{1 + x^2}{[1 + (x - \eta)^2]^2} - \frac{1}{1 + (x - \eta)^2}\right] dx \\ &\quad = \int \left[\frac{1 + (x + \eta)^2}{(1 + x^2)^2} - \frac{1}{1 + x^2}\right] dx = \int \frac{2x + \eta^2}{(1 + x^2)^2} dx = \eta^2 \int \frac{dx}{(1 + x^2)^2},\end{aligned} \tag{2.36}$$

so that

$$\Phi(P,1) \leqq c\left[\int \eta^2\,dP(\eta)\right]^2 \tag{2.37}$$

i.e.,

$$\psi(\lambda, 1) \leqq c\lambda^2 \quad \text{for} \quad \lambda \leqq \tfrac{1}{2}. \tag{2.38}$$

Combining (2.29), (2.31), and (2.38), we have finally

$$\psi(\lambda,\sigma) \leqq \begin{cases} \frac{c}{\sigma}\lambda^2 \text{ for } \lambda \leqq \frac{1}{2}\sigma^2 \\ \sigma\lambda \text{ for } \lambda \geqq \frac{1}{2}\sigma^2. \end{cases} \tag{2.39}$$

Now let $\nu^*$ be the distribution of $\int q^2\,dP(q, Y)$, i.e.,

$$\nu^*(S) = \nu\left\{y: \int \eta^2\,dP(\eta, y)\,\varepsilon S\right\}. \tag{2.40}$$

Then

$$\int d\nu(y)\int dx \frac{\left[\int \eta q\left(\frac{x-\eta}{\sigma}\right) dP(\eta, y)\right]^2}{\int q\left(\frac{x-\eta}{\sigma}\right) dP(\eta, y)} \leqq \frac{c}{\sigma}\int_0^{\frac{1}{2}\sigma^2} \lambda^2\,d\nu^*(\lambda) + \sigma\int_{\frac{1}{2}\sigma^2}^{\infty} \lambda\,d\nu^*(\lambda). \tag{2.41}$$

For any $\epsilon$ between 0 and 1, choose $\sigma_0$ so large that

$$\int_{\frac{\epsilon}{2}\sigma_0^2}^{\infty} \lambda^{3/2}\,d\nu^*(\lambda) < \epsilon. \tag{2.42}$$

Then, for $\sigma \geqq \sigma_0$,

$$\begin{aligned} \frac{1}{\sigma}\int_0^{\frac{1}{2}\sigma^2} \lambda^2\,d\nu^*(\lambda) &= \frac{1}{\sigma}\int_0^{\frac{\epsilon}{2}\sigma^2} \lambda^2\,d\nu^*(\lambda) + \frac{1}{\sigma}\int_{\frac{\epsilon}{2}\sigma^2}^{\frac{1}{2}\sigma^2} \lambda^2\,d\nu^*(\lambda) \\ &\leqq \sqrt{\frac{\epsilon}{2}}\int_0^{\frac{\epsilon}{2}\sigma^2} \lambda^{3/2}\,d\nu^*(\lambda) + \frac{1}{\sqrt{2}}\int_{\frac{\epsilon\sigma^2}{2}}^{\infty} \lambda^{3/2}\,d\nu^*(\lambda) \\ &\leqq \sqrt{\frac{\epsilon}{2}}\int_0^{\infty} \lambda^{3/2}\,d\nu^*(\lambda) + \frac{\epsilon}{\sqrt{2}}. \end{aligned} \tag{2.43}$$

$$\sigma\int_{\frac{1}{2}\sigma^2}^{\infty} \lambda\,d\nu^*(\lambda) \leqq \sqrt{2}\int_{\frac{1}{2}\sigma^2}^{\infty} \lambda^{3/2}\,d\nu^*(\lambda) \leqq \sqrt{2}\,\epsilon. \tag{2.44}$$

Thus the right-hand side of (2.41) approaches 0 as $\sigma \to \infty$, which completes the proof of (2.23), and thus the proofs of Theorems 2 and 1.

## REFERENCES

[1] E. J. G. Pitman, "The estimation of location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.

[2] M. A. GIRSHICK AND L. J. SAVAGE, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 53–73.

[3] HIROKICHI KUDO, "On minimax invariant estimates of the transformation parameter," *Natural Science Report of the Ochanomizu University*, Vol. 6 (1955), No. 1, pp. 31–73.

[4] J. KIEFER, "Invariance, minimax sequential estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 573–601.

[5] SAMUEL KARLIN, "Admissibility for estimation with quadratic loss," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 406–436.

[6] COLIN R. BLYTH, "On minimax statistical decision procedures and their admissibility," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 22–42.

[7] E. L. LEHMANN AND C. M. STEIN, "The admissibility of certain invariant statistical tests involving a translation parameter," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 473–479.

[8] CHARLES STEIN, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 1, pp. 197–206.

[9] DAVID BLACKWELL, "On the translation parameter problem for discrete variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 393–399.

[10] E. C. TITCHMARSH, *Theory of Functions*, Second edition, Clarendon Press, Oxford, 1939.

# THE USE OF SAMPLE QUASI-RANGES IN ESTIMATING POPULATION STANDARD DEVIATION

H. LEON HARTER

*Wright Air Development Center*[1]

**Summary.** The use of sample quasi-ranges in estimating the standard deviation of normal, rectangular, and exponential populations is discussed. For the normal population, the expected value and the variance of the $r$th quasi-range for samples of size $n$ are tabulated for $r = 0\ (1)\ 8$ and $n = (2r + 2)\ (1)\ 100$. The efficiency of the unbiased estimate of population standard deviation based on one sample quasi-range is tabulated for the same values of $r$, with $n = (2r + 2)\ (2)\ 50\ (5)\ 100$. Estimates based on a linear combination of two quasi-ranges are considered and a method is given for determining the weighting factor which maximizes the efficiency. The most efficient unbiased estimates based on one quasi-range for $n = 2\ (1)\ 100$ and on linear combinations of two adjacent quasi-ranges and of two quasi-ranges among those with $r < r' \leqq 8$ for $n = 4\ (1)\ 100$ are tabulated, along with their efficiencies. An example illustrates the use of these estimates. For the rectangular population, the efficient estimate of population standard deviation, which is based on the sample range, is tabulated for $n = 2\ (1)\ 100$. The bias, when estimates which assume normality are used, is tabulated for $n = 2\ (1)\ 100$ for rectangular and exponential populations.

**0. Introduction.** It is well known that, for small samples, the standard deviation of a normal population can be estimated quite efficiently from the sample range. However, the efficiency of the estimate based on the range decreases rather rapidly as the sample size increases, being less than 35% for samples of 100. There appears to be a need for substitute estimates which are reasonably efficient for moderate sample sizes, yet much simpler to compute than the efficient estimate based on the sample standard deviation. A number of authors, including Jones [9], Nair [12], [13], Godwin [6] and Sarhan and Greenberg [17] have proposed methods based on order statistics. This paper will be concerned with estimates, based on sample quasi-ranges, that satisfy these requirements quite well. Up to the present such estimates have been used relatively little, mainly because of the lack of suitable tables, though estimates based on quasi-ranges of samples of moderate size were proposed by Mosteller [11] in 1946, and estimates based on quantiles of large samples had been advocated much earlier by a number of authors, notably Edgeworth [5], Sheppard [18], and K. Pearson [15]. More recently, Benson [1] has explored further aspects of estimates of the latter type.

The $r$th quasi-range, $w_r$, of a sample of size $n$ is defined as the range of $(n - 2r)$ sample values, omitting the $r$ largest and the $r$ smallest. Symbolically,

Received August 8, 1958; revised January 14, 1959.

[1] Applied Mathematics Research Branch, Aeronautical Research Laboratory.

$w_r = x_{n-r} - x_{r+1}$, where $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ are the ordered sample values. Cadwell [2] has shown that the range, $w_0$, is the most efficient statistic of this type for estimating the standard deviation of a normal population from samples of sizes up through $n = 17$, beyond which point $w_1$ is optimum up through $n = 31$, where $w_2$ becomes better. Cadwell [2] has also proposed the use of linear combinations of quasi-ranges, while Dixon [4] has advocated the use of sums of from two to four quasi-ranges with equal weights.

Section 1 of this paper will deal with the most efficient unbiased estimates of the standard deviation of a normal population, based on one sample quasi-range and on linear combinations of two sample quasi-ranges. Section 2 will be concerned with the efficient estimate of the standard deviation of a rectangular distribution, based on the range, and with the bias of the estimates which assume normality when the population is actually rectangular or exponential.

**1. Estimates of $\sigma$ for a normal population.**

1.1. *Expected values of quasi-ranges.* In order to determine the factor by which the $r$th quasi-range, $w_r$, must be multiplied in order to obtain an unbiased estimate of the population standard deviation $\sigma$, it is necessary to know the expected value $E(w_r)$ of the $r$th quasi-range for samples of $n$ from a standard normal population, which is given by Cadwell ([2], p. 606) in terms of an integral which cannot be evaluated in closed form. Expected values of the range (to five decimal places) have been tabulated for $n = 2\ (1)\ 1000$ by Tippett [20]. Cadwell [2] has tabulated $E(w_1)$ to four decimal places for $n = 10\ (1)\ 30$. The author has computed tables of $E(w_r)$, accurate to within a unit in the sixth decimal place, for $r = 0\ (1)\ 8$ and $n = (2r + 2)\ (1)\ 100$, using the Burroughs E101 computer. The trapezoidal rule was employed for the numerical integration. The results, which are given in Table 1, agree with those obtained by Tippett and Cadwell to within a unit in the last place published by them. The values in Table 1 also agree with those found by doubling the expected values of order statistics, which have been tabulated to ten decimal places for $n = 2\ (1)\ 20$ by Teichroew [19], and rounding to six decimal places.

1.2. *Variances of quasi-ranges.* In order to determine the variances of unbiased estimates based on quasi-ranges (and hence their efficiencies), it is necessary to know the variance of the $r$th quasi-range for samples of $n$ from a standard normal population. This is given by the equation var $w_r = E(w_r^2) - [E(w_r)]^2$, where $E(w_r^2)$ can be obtained by multiplying the probability density function of $w_r$ (see Cadwell [2], p. 604) by $w_r^2$ and integrating with respect to $w_r$ between the limits 0 and $\infty$. Tippett [20] and E. S. Pearson [14] have computed approximate values of the variance of the range for a few values of $n$. Cadwell [2] has tabulated var $w_1$ to four decimal places for $n = 10\ (1)\ 30$. The variance of all quasi-ranges for samples of $n = 2\ (1)\ 20$ can be obtained quite easily from ten-decimal-place values of the variances and covariances of order statistics, which have been tabulated by Sarhan and Greenberg [17]. These tables are based on ten-decimal-place expected values of order statistics and of products of order statistics tabulated by Teichroew [19]. The author, with the assistance of Eugene

TABLE 1

Expected Value of the $r^{th}$ Quasi-Range for Samples of n from $N(\mu, 1)$

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.128379 | | | | | | | | |
| 3 | 1.692569 | | | | | | | | |
| 4 | 2.058751 | 0.594023 | | | | | | | |
| 5 | 2.325929 | 0.990038 | | | | | | | |
| 6 | 2.534413 | 1.283510 | 0.403094 | | | | | | |
| 7 | 2.704357 | 1.514749 | 0.705414 | | | | | | |
| 8 | 2.847201 | 1.704450 | 0.945645 | 0.305029 | | | | | |
| 9 | 2.970026 | 1.864595 | 1.143942 | 0.549052 | | | | | |
| 10 | 3.077505 | 2.002714 | 1.312118 | 0.751529 | 0.245336 | | | | |
| 11 | 3.172873 | 2.123833 | 1.457679 | 0.923957 | 0.449782 | | | | |
| 12 | 3.258455 | 2.231464 | 1.585676 | 1.073686 | 0.624498 | 0.205179 | | | |
| 13 | 3.335980 | 2.328154 | 1.699669 | 1.205700 | 0.776654 | 0.381047 | | | |
| 14 | 3.406763 | 2.415805 | 1.802253 | 1.323527 | 0.911132 | 0.534594 | 0.176318 | | |
| 15 | 3.471827 | 2.495870 | 1.895378 | 1.429755 | 1.031402 | 0.670592 | 0.330597 | | |
| 16 | 3.531983 | 2.569488 | 1.980542 | 1.526333 | 1.140019 | 0.792446 | 0.467503 | 0.154575 | |
| 17 | 3.587884 | 2.637564 | 2.058922 | 1.614770 | 1.238915 | 0.902667 | 0.590373 | 0.291975 | |
| 18 | 3.640064 | 2.700827 | 2.131456 | 1.696250 | 1.329589 | 1.003163 | 0.701674 | 0.415471 | 0.137605 |
| 19 | 3.688963 | 2.759877 | 2.198906 | 1.771724 | 1.413223 | 1.095415 | 0.803285 | 0.527486 | 0.261450 |
| 20 | 3.734950 | 2.815208 | 2.261896 | 1.841963 | 1.490766 | 1.180594 | 0.896664 | 0.629866 | 0.373915 |
| 21 | 3.778336 | 2.867236 | 2.320945 | 1.907604 | 1.562992 | 1.259644 | 0.982970 | 0.724051 | 0.476816 |
| 22 | 3.819385 | 2.916311 | 2.376488 | 1.969174 | 1.630538 | 1.333334 | 1.063136 | 0.811183 | 0.571570 |
| 23 | 3.858323 | 2.962731 | 2.428893 | 2.027118 | 1.693938 | 1.402301 | 1.137928 | 0.892185 | 0.659305 |
| 24 | 3.895348 | 3.006755 | 2.478476 | 2.081814 | 1.753638 | 1.467076 | 1.207975 | 0.967812 | 0.740931 |
| 25 | 3.930629 | 3.048602 | 2.525506 | 2.133585 | 1.810021 | 1.528108 | 1.273807 | 1.038691 | 0.817195 |

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 26 | 3.964316 | 3.088468 | 2.570219 | 2.182707 | 1.863411 | 1.585779 | 1.335872 | 1.105347 | 0.888717 |
| 27 | 3.996539 | 3.126520 | 2.612819 | 2.229423 | 1.914092 | 1.640416 | 1.394549 | 1.168222 | 0.956017 |
| 28 | 4.027414 | 3.162907 | 2.653484 | 2.273942 | 1.962307 | 1.692302 | 1.450167 | 1.227697 | 1.019535 |
| 29 | 4.057044 | 3.197761 | 2.692372 | 2.316449 | 2.008271 | 1.741683 | 1.503007 | 1.284097 | 1.079647 |
| 30 | 4.085522 | 3.231200 | 2.729624 | 2.357108 | 2.052170 | 1.788775 | 1.553316 | 1.337704 | 1.136678 |
| 31 | 4.112928 | 3.263326 | 2.765363 | 2.396061 | 2.094171 | 1.833766 | 1.601311 | 1.388764 | 1.190907 |
| 32 | 4.139338 | 3.294235 | 2.799700 | 2.433439 | 2.134420 | 1.876825 | 1.647180 | 1.437492 | 1.242580 |
| 33 | 4.164817 | 3.324009 | 2.832734 | 2.469355 | 2.173049 | 1.918099 | 1.691092 | 1.484078 | 1.291910 |
| 34 | 4.189425 | 3.352725 | 2.864556 | 2.503912 | 2.210174 | 1.957721 | 1.733196 | 1.528690 | 1.339088 |
| 35 | 4.213219 | 3.380451 | 2.895245 | 2.537204 | 2.245901 | 1.995809 | 1.773626 | 1.571478 | 1.384281 |
| 36 | 4.236247 | 3.407249 | 2.924875 | 2.569314 | 2.280326 | 2.032471 | 1.812500 | 1.612575 | 1.427639 |
| 37 | 4.258554 | 3.433177 | 2.953513 | 2.600317 | 2.313532 | 2.067802 | 1.849927 | 1.652101 | 1.469294 |
| 38 | 4.280183 | 3.458286 | 2.981218 | 2.630284 | 2.345600 | 2.101890 | 1.886002 | 1.690164 | 1.509367 |
| 39 | 4.301171 | 3.482623 | 3.008047 | 2.659277 | 2.376598 | 2.134813 | 1.920814 | 1.726860 | 1.547965 |
| 40 | 4.321554 | 3.506233 | 3.034049 | 2.687353 | 2.406591 | 2.166643 | 1.954443 | 1.762279 | 1.585186 |
| 41 | 4.341364 | 3.529154 | 3.059272 | 2.714565 | 2.435639 | 2.197445 | 1.986961 | 1.796500 | 1.621118 |
| 42 | 4.360631 | 3.551424 | 3.083756 | 2.740962 | 2.463796 | 2.227280 | 2.018434 | 1.829596 | 1.655842 |
| 43 | 4.379382 | 3.573076 | 3.107544 | 2.766588 | 2.491111 | 2.256203 | 2.048923 | 1.861634 | 1.689430 |
| 44 | 4.397644 | 3.594143 | 3.130670 | 2.791484 | 2.517629 | 2.284264 | 2.078483 | 1.892675 | 1.721950 |
| 45 | 4.415439 | 3.614654 | 3.153169 | 2.815688 | 2.543394 | 2.311510 | 2.107166 | 1.922775 | 1.753463 |
| 46 | 4.432790 | 3.634635 | 3.175071 | 2.839235 | 2.568444 | 2.337984 | 2.135019 | 1.951985 | 1.784025 |
| 47 | 4.449718 | 3.654111 | 3.196406 | 2.862158 | 2.592816 | 2.363725 | 2.162086 | 1.980354 | 1.813688 |
| 48 | 4.466242 | 3.673108 | 3.217201 | 2.884486 | 2.616542 | 2.388771 | 2.188406 | 2.007924 | 1.842500 |
| 49 | 4.482379 | 3.691645 | 3.237481 | 2.906249 | 2.639654 | 2.413155 | 2.214017 | 2.034738 | 1.870505 |
| 50 | 4.498147 | 3.709744 | 3.257268 | 2.927472 | 2.662181 | 2.436910 | 2.238954 | 2.060832 | 1.897744 |

**TABLE 1 (continued)**

**Expected Value of the $r^{th}$ Quasi-Range for Samples of $n$ from $N(\mu, 1)$**

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 51 | 4.513562 | 3.727424 | 3.276586 | 2.948181 | 2.684150 | 2.460065 | 2.263249 | 2.086242 | 1.924256 |
| 52 | 4.528637 | 3.744702 | 3.295455 | 2.968397 | 2.705587 | 2.482647 | 2.286933 | 2.110999 | 1.950074 |
| 53 | 4.543388 | 3.761597 | 3.313894 | 2.988142 | 2.726514 | 2.504683 | 2.310032 | 2.135136 | 1.975233 |
| 54 | 4.557827 | 3.778123 | 3.331921 | 3.007438 | 2.746955 | 2.526197 | 2.332574 | 2.158679 | 1.999762 |
| 55 | 4.571967 | 3.794295 | 3.349553 | 3.026301 | 2.766929 | 2.547210 | 2.354583 | 2.181655 | 2.023691 |
| 56 | 4.585818 | 3.810128 | 3.366805 | 3.044750 | 2.786457 | 2.567746 | 2.376082 | 2.204090 | 2.047045 |
| 57 | 4.599393 | 3.825635 | 3.383695 | 3.062803 | 2.805556 | 2.587823 | 2.397093 | 2.226007 | 2.069851 |
| 58 | 4.612701 | 3.840828 | 3.400234 | 3.080474 | 2.824244 | 2.607460 | 2.417635 | 2.247427 | 2.092132 |
| 59 | 4.625752 | 3.855719 | 3.416437 | 3.097778 | 2.842538 | 2.626675 | 2.437728 | 2.268371 | 2.113909 |
| 60 | 4.638556 | 3.870319 | 3.432316 | 3.114730 | 2.860452 | 2.645484 | 2.457390 | 2.288858 | 2.135204 |
| 61 | 4.651122 | 3.884639 | 3.447884 | 3.131343 | 2.878001 | 2.663904 | 2.476638 | 2.308906 | 2.156036 |
| 62 | 4.663457 | 3.898688 | 3.463151 | 3.147629 | 2.895199 | 2.681948 | 2.495488 | 2.328534 | 2.176423 |
| 63 | 4.675569 | 3.912477 | 3.478128 | 3.163599 | 2.912058 | 2.699632 | 2.513954 | 2.347756 | 2.196382 |
| 64 | 4.687467 | 3.926014 | 3.492827 | 3.179267 | 2.928591 | 2.716968 | 2.532052 | 2.366588 | 2.215931 |
| 65 | 4.699157 | 3.939308 | 3.507255 | 3.194641 | 2.944809 | 2.733968 | 2.549794 | 2.385044 | 2.235084 |
| 66 | 4.710646 | 3.952367 | 3.521423 | 3.209732 | 2.960724 | 2.750646 | 2.567194 | 2.403139 | 2.253856 |
| 67 | 4.721941 | 3.965199 | 3.535339 | 3.224550 | 2.976347 | 2.767011 | 2.584263 | 2.420886 | 2.272261 |
| 68 | 4.733047 | 3.977811 | 3.549011 | 3.239104 | 2.991685 | 2.783075 | 2.601013 | 2.438295 | 2.290312 |
| 69 | 4.743971 | 3.990210 | 3.562448 | 3.253403 | 3.006751 | 2.798849 | 2.617456 | 2.455380 | 2.308021 |
| 70 | 4.754718 | 4.002402 | 3.575656 | 3.267455 | 3.021552 | 2.814341 | 2.633601 | 2.472152 | 2.325401 |
| 71 | 4.765294 | 4.014395 | 3.588644 | 3.281267 | 3.036096 | 2.829561 | 2.649458 | 2.488620 | 2.342462 |
| 72 | 4.775704 | 4.026195 | 3.601418 | 3.294848 | 3.050393 | 2.844517 | 2.665037 | 2.504796 | 2.359216 |
| 73 | 4.785953 | 4.037806 | 3.613984 | 3.308204 | 3.064450 | 2.859219 | 2.680347 | 2.520688 | 2.375673 |
| 74 | 4.796045 | 4.049236 | 3.626349 | 3.321343 | 3.078275 | 2.873674 | 2.695396 | 2.536305 | 2.391841 |
| 75 | 4.805985 | 4.060488 | 3.638519 | 3.334271 | 3.091874 | 2.887890 | 2.710192 | 2.551658 | 2.407731 |
| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
| 76 | 4.815777 | 4.071569 | 3.650499 | 3.346994 | 3.105254 | 2.901874 | 2.724744 | 2.566752 | 2.423351 |
| 77 | 4.825426 | 4.082483 | 3.662296 | 3.359519 | 3.118422 | 2.915633 | 2.739059 | 2.581598 | 2.438709 |
| 78 | 4.834935 | 4.093235 | 3.673914 | 3.371850 | 3.131384 | 2.929174 | 2.753143 | 2.596201 | 2.453815 |
| 79 | 4.844308 | 4.103829 | 3.685358 | 3.383994 | 3.144146 | 2.942503 | 2.767004 | 2.610571 | 2.468674 |
| 80 | 4.853549 | 4.114270 | 3.696633 | 3.395957 | 3.156714 | 2.955626 | 2.780649 | 2.624712 | 2.483296 |
| 81 | 4.862661 | 4.124561 | 3.707745 | 3.407742 | 3.169094 | 2.968550 | 2.794083 | 2.638633 | 2.497686 |
| 82 | 4.871648 | 4.134708 | 3.718696 | 3.419355 | 3.181289 | 2.981279 | 2.807312 | 2.652339 | 2.511851 |
| 83 | 4.880513 | 4.144713 | 3.729492 | 3.430800 | 3.193307 | 2.993819 | 2.820343 | 2.665836 | 2.525799 |
| 84 | 4.889259 | 4.154581 | 3.740137 | 3.442082 | 3.205150 | 3.006176 | 2.833180 | 2.679131 | 2.539534 |
| 85 | 4.897890 | 4.164315 | 3.750635 | 3.453206 | 3.216825 | 3.018355 | 2.845830 | 2.692229 | 2.553063 |
| 86 | 4.906407 | 4.173918 | 3.760988 | 3.464176 | 3.228335 | 3.030359 | 2.858296 | 2.705135 | 2.566393 |
| 87 | 4.914814 | 4.183393 | 3.771203 | 3.474994 | 3.239685 | 3.042194 | 2.870585 | 2.717855 | 2.579527 |
| 88 | 4.923114 | 4.192745 | 3.781280 | 3.485666 | 3.250879 | 3.053864 | 2.882700 | 2.730393 | 2.592471 |
| 89 | 4.931308 | 4.201975 | 3.791225 | 3.496196 | 3.261921 | 3.065374 | 2.894647 | 2.742754 | 2.605232 |
| 90 | 4.939401 | 4.211087 | 3.801040 | 3.506585 | 3.272815 | 3.076727 | 2.906429 | 2.754944 | 2.617812 |
| 91 | 4.947393 | 4.220084 | 3.810729 | 3.516839 | 3.283565 | 3.087928 | 2.918051 | 2.766965 | 2.630218 |
| 92 | 4.955288 | 4.228968 | 3.820294 | 3.526960 | 3.294173 | 3.098980 | 2.929517 | 2.778824 | 2.642452 |
| 93 | 4.963087 | 4.237743 | 3.829739 | 3.536952 | 3.304644 | 3.109887 | 2.940830 | 2.790523 | 2.654521 |
| 94 | 4.970794 | 4.246410 | 3.839066 | 3.546818 | 3.314980 | 3.120652 | 2.951995 | 2.802066 | 2.666428 |
| 95 | 4.978409 | 4.254972 | 3.848278 | 3.556560 | 3.325186 | 3.131279 | 2.963015 | 2.813458 | 2.678176 |
| 96 | 4.985935 | 4.263431 | 3.857378 | 3.566181 | 3.335264 | 3.141772 | 2.973894 | 2.824702 | 2.689771 |
| 97 | 4.993374 | 4.271790 | 3.866368 | 3.575685 | 3.345216 | 3.152132 | 2.984634 | 2.835802 | 2.701215 |
| 98 | 5.000728 | 4.280051 | 3.875251 | 3.585074 | 3.355047 | 3.162364 | 2.995240 | 2.846761 | 2.712512 |
| 99 | 5.007998 | 4.288217 | 3.884029 | 3.594350 | 3.364759 | 3.172471 | 3.005713 | 2.857582 | 2.723666 |
| 100 | 5.015187 | 4.296289 | 3.892705 | 3.603517 | 3.374354 | 3.182455 | 3.016059 | 2.868269 | 2.734680 |

H. Guthrie, has computed tables of var $w_r$, accurate to within a unit in the fifth decimal place, for $r = 0\ (1)\ 8$ and $n = (2r + 2)\ (1)\ 100$, using the Univac Scientific (ERA 1103) computer. Since Cadwell's expression for the probability density function of $w_r$ involves an integral with respect to another variable $x$, it was necessary to integrate numerically with respect to both $w_r$ and $x$. A seven-point integration formula was employed for a few cases where the trapezoidal rule did not give sufficient accuracy for the integral with respect to $w_r$, while the trapezoidal rule was used for all cases in the integration with respect to $x$. The variances, which are shown in Table 2, agree with those obtained by Tippett, Pearson, and Cadwell to within a unit in the last place published by them. The values in Table 2 also agree, to within a unit in the fifth decimal place, with results computed from the Sarhan and Greenberg table of variances and covariances of order statistics for $n = 2\ (1)\ 20$.

1.3. *Covariance of two quasi-ranges.* In order to determine the variances of unbiased estimates based on linear combinations of two quasi-ranges (and hence their efficiencies), it is necessary to know not only the variances of the two quasi-ranges but also their covariance. The covariance of the $r$th and $r'$th quasi-ranges for samples of size $n$ is given by $\text{cov}\,(w_r, w_{r'}) = E(w_r w_{r'}) - E(w_r)E(w_{r'})$, in which $E(w_r w_{r'}) = 2[E(x_{r+1}x_{r'+1}) - E(x_{r+1}x_{n-r'})]$. Wilks ([21], p. 20) has given an expression for the joint probability density function of the $k$th and $k'$th order statistics. Godwin [7] has tabulated (to five decimal places) the covariances of all order statistics for samples of $n = 2\ (1)\ 10$. The more extensive and more precise tables of Sarhan and Greenberg [17] were mentioned in the preceding paragraph. The author, again with the assistance of Eugene H. Guthrie, undertook the task of computing $\text{cov}\,(w_r, w_{r'})$ for $0 \leqq r < r' \leqq 8$ and $n = (2r' + 2)\ (1)\ 100$, accurate to within a unit in the fifth decimal place, using the Univac Scientific (ERA 1103) computer. Finding that the complete tabulation required too much machine time, he decided to limit the computations to those values required to determine the most efficient estimates of the population standard deviation based on linear combinations of two adjacent quasi-ranges and of two quasi-ranges among those with $r < r' \leqq 8$ for $n = 4\ (1)\ 100$, together with the numerical values of the efficiencies of these estimates. Wilks' expression for the joint probability density function of $x_k$ and $x_{k'}$ was first integrated with respect to $x_{k'}$ between the limits $x_k$ and $\infty$ by using a seven-point integration formula; the result was then integrated with respect to $x_k$ between the limits $-\infty$ and $\infty$ by employing the trapezoidal rule. For $n = 4\ (1)\ 20$, the expected values of products of order statistics computed in this manner agree with Teichroew's values to within a unit in the sixth decimal place and the covariances of quasi-ranges computed from these results agree to within a unit in the fifth decimal place with those computed from the covariances of order statistics tabulated by Sarhan and Greenberg.

1.4. *Unbiased estimates of population standard deivation.* The minimum variance unbiased estimate (which will hereafter be called the efficient estimate) of population standard deviation $\sigma$ is the one based on the sample standard deviation $s$,

and given by the equation $\hat{\sigma} = s/c_2$, where $s = [\sum (x - \bar{x})^2/(n - 1)]^{\frac{1}{2}}$ and $c_2 = [2/(n - 1)]^{\frac{1}{2}}\Gamma(n/2)/\Gamma[(n - 1)/2]$. The unbiased estimate of $\sigma$ based on one sample quasi-range is given by the equation $\bar{\sigma}_r = w_r/E(w_r)$ while the unbiased estimate of $\sigma$ based on a linear combination of two sample quasi-ranges is given by the equation

(1) $$\bar{\sigma}_{r,r'} = \frac{w_r + \lambda_{r,r'} w_{r'}}{E(w_r) + \lambda_{r,r'} E(w_{r'})},$$

TABLE 2

Variance of the $r^{th}$ Quasi-Range for Samples of n from $N(\mu, 1)$

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | .72676 | | | | | | | | |
| 3 | .78920 | | | | | | | | |
| 4 | .77406 | .24902 | | | | | | | |
| 5 | .74664 | .32315 | | | | | | | |
| 6 | .71917 | .34734 | .12588 | | | | | | |
| 7 | .69423 | .35350 | .18023 | | | | | | |
| 8 | .67212 | .35231 | .20590 | .07602 | | | | | |
| 9 | .65259 | .34796 | .21829 | .11578 | | | | | |
| 10 | .63529 | .34231 | .22394 | .13799 | .05091 | | | | |
| 11 | .61986 | .33619 | .22594 | .15081 | .08089 | | | | |
| 12 | .60603 | .33003 | .22589 | .15830 | .09943 | .03649 | | | |
| 13 | .59354 | .32402 | .22466 | .16256 | .11125 | .05980 | | | |
| 14 | .58220 | .31827 | .22275 | .16481 | .11892 | .07525 | .02744 | | |
| 15 | .57185 | .31280 | .22045 | .16576 | .12391 | .08577 | .04604 | | |
| 16 | .56235 | .30763 | .21796 | .16588 | .12713 | .09305 | .05902 | .02138 | |
| 17 | .55361 | .30275 | .21537 | .16544 | .12913 | .09814 | .06828 | .03655 | |
| 18 | .54551 | .29816 | .21277 | .16462 | .13028 | .10171 | .07499 | .04757 | 01714 |
| 19 | .53799 | .29382 | .21019 | .16356 | .13084 | .10419 | .07991 | .05572 | .02973 |
| 20 | .53098 | .28973 | .20765 | .16234 | .13097 | .10588 | .08353 | .06183 | .03918 |
| 21 | .52442 | .28586 | .20518 | .16101 | .13079 | .10700 | .08620 | .06646 | .04637 |
| 22 | .51827 | .28220 | .20279 | .15963 | .13040 | .10768 | .08817 | .07000 | .05191 |
| 23 | .51249 | .27874 | .20048 | .15822 | .12984 | .10805 | .08959 | .07271 | .05622 |
| 24 | .50703 | .27545 | .19825 | .15679 | .12917 | .10817 | .09061 | .07478 | .05960 |
| 25 | .50188 | .27233 | .19611 | .15537 | .12841 | .10810 | .09131 | .07637 | .06226 |
| 26 | .49699 | .26936 | .19404 | .15396 | .12760 | .10789 | .09175 | .07757 | .06436 |
| 27 | .49236 | .26653 | .19205 | .15258 | .12675 | .10757 | .09201 | .07848 | .06602 |
| 28 | .48796 | .26383 | .19014 | .15122 | .12587 | .10717 | .09211 | .07914 | .06733 |
| 29 | .48377 | .26126 | .18830 | .14989 | .12498 | .10671 | .09209 | .07961 | .06836 |
| 30 | .47977 | .25879 | .18652 | .14859 | .12409 | .10619 | .09198 | .07992 | .06916 |
| 31 | .47595 | .25644 | .18481 | .14732 | .12319 | .10565 | .09178 | .08011 | .06977 |
| 32 | .47229 | .25417 | .18316 | .14609 | .12230 | .10507 | .09153 | .08020 | .07023 |
| 33 | .46879 | .25201 | .18158 | .14489 | .12141 | .10448 | .09123 | .08020 | .07056 |
| 34 | .46544 | .24992 | .18004 | .14372 | .12054 | .10388 | .09089 | .08014 | .07079 |
| 35 | .46221 | .24792 | .17857 | .14259 | .11969 | .10327 | .09052 | .08002 | .07094 |

TABLE 2 (continued)

Variance of the $r^{th}$ Quasi-Range for Samples of n from $N(\mu, 1)$

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 36 | .45910 | .24599 | .17714 | .14148 | .11884 | .10265 | .09013 | .07985 | .07101 |
| 37 | .45612 | .24413 | .17576 | .14042 | .11801 | .10204 | .08972 | .07965 | .07103 |
| 38 | .45325 | .24234 | .17443 | .13937 | .11719 | .10142 | .08930 | .07941 | .07099 |
| 39 | .45048 | .24062 | .17313 | .13836 | .11640 | .10081 | .08886 | .07916 | .07091 |
| 40 | .44780 | .23894 | .17188 | .13738 | .11562 | .10021 | .08842 | .07888 | .07080 |
| 41 | .44521 | .23733 | .17067 | .13642 | .11486 | .09961 | .08797 | .07858 | .07066 |
| 42 | .44271 | .23577 | .16950 | .13549 | .11411 | .09902 | .08753 | .07827 | .07049 |
| 43 | .44029 | .23426 | .16837 | .13458 | .11337 | .09844 | .08708 | .07796 | .07031 |
| 44 | .43794 | .23280 | .16726 | .13370 | .11266 | .09786 | .08663 | .07763 | .07010 |
| 45 | .43567 | .23137 | .16619 | .13284 | .11196 | .09730 | .08619 | .07730 | .06988 |
| 46 | .43347 | .23000 | .16515 | .13201 | .11128 | .09674 | .08574 | .07696 | .06965 |
| 47 | .43133 | .22867 | .16414 | .13119 | .11061 | .09619 | .08530 | .07662 | .06942 |
| 48 | .42925 | .22736 | .16316 | .13040 | .10996 | .09565 | .08486 | .07629 | .06917 |
| 49 | .42724 | .22611 | .16220 | .12963 | .10932 | .09512 | .08443 | .07594 | .06892 |
| 50 | .42527 | .22488 | .16127 | .12888 | .10869 | .09460 | .08401 | .07560 | .06866 |
| 51 | .42336 | .22369 | .16036 | .12814 | .10808 | .09409 | .08359 | .07526 | .06840 |
| 52 | .42151 | .22253 | .15948 | .12742 | .10748 | .09359 | .08317 | .07492 | .06813 |
| 53 | .41970 | .22140 | .15862 | .12673 | .10690 | .09310 | .08276 | .07458 | .06787 |
| 54 | .41794 | .22030 | .15778 | .12604 | .10633 | .09262 | .08235 | .07425 | .06760 |
| 55 | .41621 | .21923 | .15696 | .12537 | .10577 | .09215 | .08195 | .07392 | .06733 |
| 56 | .41454 | .21819 | .15616 | .12473 | .10522 | .09168 | .08156 | .07359 | .06707 |
| 57 | .41291 | .21716 | .15538 | .12408 | .10469 | .09122 | .08117 | .07327 | .06680 |
| 58 | .41131 | .21617 | .15462 | .12346 | .10417 | .09078 | .08079 | .07294 | .06653 |
| 59 | .40976 | .21519 | .15388 | .12286 | .10365 | .09034 | .08042 | .07262 | .06627 |
| 60 | .40823 | .21424 | .15315 | .12226 | .10315 | .08991 | .08005 | .07231 | .06600 |
| 61 | .40674 | .21332 | .15244 | .12167 | .10265 | .08948 | .07968 | .07200 | .06574 |
| 62 | .40528 | .21241 | .15175 | .12110 | .10217 | .08907 | .07932 | .07169 | .06547 |
| 63 | .40387 | .21153 | .15107 | .12055 | .10170 | .08866 | .07897 | .07138 | .06522 |
| 64 | .40247 | .21066 | .15040 | .12000 | .10123 | .08826 | .07862 | .07108 | .06496 |
| 65 | .40111 | .20981 | .14975 | .11946 | .10078 | .08787 | .07828 | .07079 | .06470 |
| 66 | .39978 | .20898 | .14911 | .11894 | .10033 | .08748 | .07794 | .07050 | .06445 |
| 67 | .39847 | .20816 | .14849 | .11843 | .09989 | .08710 | .07761 | .07021 | .06420 |
| 68 | .39719 | .20737 | .14788 | .11792 | .09947 | .08673 | .07729 | .06993 | .06396 |
| 69 | .39594 | .20659 | .14728 | .11743 | .09904 | .08636 | .07697 | .06965 | .06371 |
| 70 | .39471 | .20583 | .14669 | .11694 | .09863 | .08600 | .07665 | .06937 | .06347 |

**TABLE 2 (continued)**

Variance of the $r^{th}$ Quasi-Range for Samples of n from $N(\mu, 1)$

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 71 | .39351 | .20508 | .14612 | .11647 | .09822 | .08565 | .07634 | .06910 | .06324 |
| 72 | .39232 | .20434 | .14555 | .11600 | .09783 | .08531 | .07604 | .06882 | .06300 |
| 73 | .39116 | .20362 | .14500 | .11554 | .09743 | .08496 | .07574 | .06856 | .06276 |
| 74 | .39002 | .20291 | .14446 | .11509 | .09704 | .08463 | .07544 | .06830 | .06253 |
| 75 | .38890 | .20222 | .14392 | .11465 | .09667 | .08430 | .07515 | .06804 | .06230 |
| 76 | .38780 | .20154 | .14340 | .11422 | .09630 | .08397 | .07486 | .06779 | .06208 |
| 77 | .38672 | .20087 | .14289 | .11379 | .09593 | .08365 | .07458 | .06753 | .06186 |
| 78 | .38566 | .20021 | .14238 | .11337 | .09557 | .08334 | .07430 | .06729 | .06164 |
| 79 | .38463 | .19957 | .14188 | .11296 | .09522 | .08303 | .07403 | .06704 | .06142 |
| 80 | .38360 | .19894 | .14140 | .11255 | .09487 | .08273 | .07376 | .06680 | .06120 |
| 81 | .38260 | .19832 | .14092 | .11216 | .09453 | .08242 | .07349 | .06656 | .06099 |
| 82 | .38161 | .19770 | .14045 | .11177 | .09420 | .08213 | .07323 | .06633 | .06079 |
| 83 | .38064 | .19711 | .13999 | .11138 | .09386 | .08184 | .07297 | .06610 | .06058 |
| 84 | .37969 | .19651 | .13953 | .11101 | .09354 | .08156 | .07272 | .06588 | .06037 |
| 85 | .37874 | .19593 | .13908 | .11064 | .09322 | .08127 | .07247 | .06565 | .06018 |
| 86 | .37782 | .19536 | .13865 | .11027 | .09290 | .08100 | .07223 | .06543 | .05997 |
| 87 | .37691 | .19480 | .13820 | .10991 | .09259 | .08072 | .07198 | .06521 | .05978 |
| 88 | .37601 | .19424 | .13778 | .10956 | .09229 | .08046 | .07174 | .06499 | .05959 |
| 89 | .37513 | .19370 | .13736 | .10920 | .09199 | .08019 | .07150 | .06479 | .05939 |
| 90 | .37426 | .19316 | .13695 | .10886 | .09169 | .07993 | .07127 | .06457 | .05920 |
| 91 | .37340 | .19264 | .13654 | .10852 | .09139 | .07967 | .07104 | .06437 | .05901 |
| 92 | .37256 | .19212 | .13614 | .10819 | .09111 | .07941 | .07081 | .06416 | .05883 |
| 93 | .37173 | .19160 | .13575 | .10786 | .09082 | .07916 | .07059 | .06396 | .05865 |
| 94 | .37091 | .19109 | .13536 | .10753 | .09055 | .07892 | .07037 | .06376 | .05847 |
| 95 | .37010 | .19060 | .13498 | .10721 | .09027 | .07868 | .07015 | .06357 | .05829 |
| 96 | .36931 | .19011 | .13460 | .10691 | .08999 | .07843 | .06994 | .06337 | .05812 |
| 97 | .36852 | .18963 | .13423 | .10659 | .08973 | .07820 | .06973 | .06318 | .05794 |
| 98 | .36774 | .18916 | .13386 | .10629 | .08946 | .07797 | .06952 | .06299 | .05777 |
| 99 | .36699 | .18868 | .13350 | .10599 | .08920 | .07773 | .06931 | .06281 | .05760 |
| 100 | .36624 | .18822 | .13314 | .10569 | .08894 | .07750 | .06911 | .06262 | .05744 |

where $\lambda_{r,r'}$ is a weighting factor. In the expressions for $\tilde{\sigma}_r$ and $\tilde{\sigma}_{r,r'}$, the expected values are understood to be those for samples drawn from $N(0, 1)$, the standard normal population.

1.5. *Efficiency of unbiased estimates of* $\sigma$. The efficiency of the efficient estimate $\hat{\sigma}$ is by definition 1 (100%). The efficiency of a substitute estimate is defined as the ratio of the variance of the efficient estimate to the variance of the substitute estimate. Thus the efficiency of $\tilde{\sigma}_r$ is given by Eff $\tilde{\sigma}_r$ = var $\hat{\sigma}$/var $\tilde{\sigma}_r$, while the efficiency of $\tilde{\sigma}_{r,r'}$ is given by Eff $\tilde{\sigma}_{r,r'}$ = var $\hat{\sigma}$/var $\tilde{\sigma}_{r,r'}$, where var $\hat{\sigma} = [(1 - c_2^2)/c_2^2]\sigma^2$. By varying the weighting factor $\lambda_{r,r'}$, one may obtain a one-parameter family of unbiased estimates $\tilde{\sigma}_{r,r'}$. However, there is just one

TABLE 3

**Efficiency (Percent) of Estimate of Population Standard Deviation Based on the $r^{th}$ Quasi-Range for Samples of n from $N(\mu, \sigma^2)$**

| n | r = 0 | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 | r = 8 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 100.00 | | | | | | | | |
| 4 | 97.52 | 25.24 | | | | | | | |
| 6 | 93.30 | 49.55 | 13.48 | | | | | | |
| 8 | 89.00 | 60.84 | 32.05 | 9.03 | | | | | |
| 10 | 84.99 | 66.80 | 43.83 | 23.33 | 6.74 | | | | |
| 12 | 81.36 | 70.07 | 51.69 | 33.82 | 18.21 | 5.36 | | | |
| 14 | 78.09 | 71.83 | 57.12 | 41.64 | 27.34 | 14.88 | 4.44 | | |
| 16 | 75.13 | 72.69 | 60.95 | 47.57 | 34.63 | 22.86 | 12.54 | 3.78 | |
| 18 | 72.46 | 72.98 | 63.70 | 52.14 | 40.48 | 29.52 | 19.58 | 10.82 | 3.30 |
| 20 | 70.02 | 72.91 | 65.67 | 55.70 | 45.23 | 35.09 | 25.65 | 17.10 | 9.51 |
| 22 | 67.80 | 72.59 | 67.08 | 58.51 | 49.11 | 39.76 | 30.88 | 22.64 | 15.16 |
| 24 | 65.75 | 72.11 | 68.07 | 60.73 | 52.31 | 43.72 | 35.38 | 27.52 | 20.24 |
| 26 | 63.86 | 71.52 | 68.76 | 62.49 | 54.96 | 47.07 | 39.28 | 31.81 | 24.78 |
| 28 | 62.12 | 70.86 | 69.20 | 63.90 | 57.17 | 49.94 | 42.66 | 35.59 | 28.85 |
| 30 | 60.49 | 70.15 | 69.46 | 65.01 | 59.01 | 52.39 | 45.61 | 38.93 | 32.49 |
| 32 | 58.98 | 69.41 | 69.57 | 65.90 | 60.56 | 54.50 | 48.19 | 41.89 | 35.74 |
| 34 | 57.56 | 68.66 | 69.57 | 66.59 | 61.86 | 56.32 | 50.45 | 44.51 | 38.66 |
| 36 | 56.23 | 67.89 | 69.48 | 67.12 | 62.95 | 57.89 | 52.44 | 46.85 | 41.29 |
| 38 | 54.98 | 67.13 | 69.32 | 67.53 | 63.86 | 59.26 | 54.19 | 48.93 | 43.66 |
| 40 | 53.81 | 66.38 | 69.10 | 67.82 | 64.63 | 60.44 | 55.74 | 50.80 | 45.79 |
| 42 | 52.70 | 65.63 | 68.83 | 68.03 | 65.27 | 61.46 | 57.11 | 52.47 | 47.72 |
| 44 | 51.64 | 64.89 | 68.53 | 68.16 | 65.80 | 62.35 | 58.32 | 53.97 | 49.47 |
| 46 | 50.64 | 64.17 | 68.20 | 68.22 | 66.23 | 63.13 | 59.39 | 55.31 | 51.05 |
| 48 | 49.70 | 63.46 | 67.84 | 68.23 | 66.59 | 63.80 | 60.35 | 56.52 | 52.49 |
| 50 | 48.79 | 62.76 | 67.47 | 68.20 | 66.87 | 64.38 | 61.20 | 57.61 | 53.79 |
| 55 | 46.71 | 61.08 | 66.49 | 67.95 | 67.33 | 65.50 | 62.92 | 59.89 | 56.58 |
| 60 | 44.85 | 59.50 | 65.46 | 67.53 | 67.51 | 66.25 | 64.20 | 61.66 | 58.79 |
| 65 | 43.18 | 58.01 | 64.42 | 67.00 | 67.49 | 66.71 | 65.14 | 63.02 | 60.55 |
| 70 | 41.65 | 56.60 | 63.38 | 66.39 | 67.32 | 66.98 | 65.80 | 64.07 | 61.96 |
| 75 | 40.26 | 55.27 | 62.36 | 65.74 | 67.04 | 67.07 | 66.26 | 64.88 | 63.08 |
| 80 | 38.99 | 54.02 | 61.36 | 65.05 | 66.69 | 67.04 | 66.56 | 65.47 | 63.97 |
| 85 | 37.81 | 52.84 | 60.38 | 64.35 | 66.27 | 66.92 | 66.72 | 65.91 | 64.67 |
| 90 | 36.73 | 51.72 | 59.43 | 63.63 | 65.81 | 66.72 | 66.77 | 66.22 | 65.21 |
| 95 | 35.72 | 50.66 | 58.51 | 62.92 | 65.33 | 66.46 | 66.74 | 66.41 | 65.62 |
| 100 | 34.77 | 49.65 | 57.62 | 62.21 | 64.82 | 66.16 | 66.65 | 66.52 | 65.93 |

value of $\lambda_{r,r'}$ which minimizes $V_{r,r'} = \operatorname{var} \tilde{\sigma}_{r,r'}$, and hence maximizes Eff $\tilde{\sigma}_{r,r'}$. This value of $\lambda_{r,r'}$ which maximizes the efficiency of the estimate may be obtained by setting $dV_{r,r'}/d\lambda_{r,r'} = 0$ and solving for $\lambda_{r,r'}$, which yields

$$\lambda_{r,r'} = \frac{E(w_r) \operatorname{var} w_{r'} - E(w_{r'}) \operatorname{cov}(w_r, w_{r'})}{E(w_{r'}) \operatorname{var} w_r - E(w_r) \operatorname{cov}(w_r, w_{r'})}. \tag{2}$$

Table 3 shows the efficiency of estimates based on $w_r$ for $r = 0\,(1)\,8$ and $n = (2r + 2)\,(2)\,50\,(5)\,100$, accurate to within 0.01%. Table 4 gives the

TABLE 4

Most Efficient Unbiased Estimates of Standard Deviation of Normal Population

| Sample size, n | Based on one quasi-range | | Based on a linear combination of two adjacent quasi-ranges | | Based on a linear combination of two quasi-ranges among those with $r < r' \le 8$ | | Efficient Estimate |
|---|---|---|---|---|---|---|---|
| | Estimate | Eff(%) | Estimate | Eff(%) | Estimate | Eff(%) | |
| 2 | $.886227\ w_o$ | 100.00 | | | | | $s/.797885$ |
| 3 | $.590818\ w_o$ | 99.19 | | | | | $s/.886227$ |
| 4 | $.485731\ w_o$ | 97.52 | $.45394\ (w_o + 0.2427\ w_1)$ | 98.92 | $.45394\ (w_o + 0.2427\ w_1)$ | 98.92 | $s/.921318$ |
| 5 | $.429936\ w_o$ | 95.48 | $.37238\ (w_o + 0.3631\ w_1)$ | 98.84 | $.37238\ (w_o + 0.3631\ w_1)$ | 98.84 | $s/.939986$ |
| 6 | $.394569\ w_o$ | 93.30 | $.31803\ (w_o + 0.4752\ w_1)$ | 98.66 | $.31803\ (w_o + 0.4752\ w_1)$ | 98.66 | $s/.951533$ |
| 7 | $.369774\ w_o$ | 91.12 | $.27922\ (w_o + 0.5790\ w_1)$ | 98.32 | $.27922\ (w_o + 0.5790\ w_1)$ | 98.32 | $s/.959369$ |
| 8 | $.351222\ w_o$ | 89.00 | $.25010\ (w_o + 0.6754\ w_1)$ | 97.84 | $.25010\ (w_o + 0.6754\ w_1)$ | 97.84 | $s/.965030$ |
| 9 | $.336697\ w_o$ | 86.95 | $.22745\ (w_o + 0.7651\ w_1)$ | 97.23 | $.22745\ (w_o + 0.7651\ w_1)$ | 97.23 | $s/.969311$ |
| 10 | $.324939\ w_o$ | 84.99 | $.20931\ (w_o + 0.8489\ w_1)$ | 96.54 | $.20931\ (w_o + 0.8489\ w_1)$ | 96.54 | $s/.972659$ |
| 11 | $.315172\ w_o$ | 83.13 | $.19444\ (w_o + 0.9276\ w_1)$ | 95.78 | $.19444\ (w_o + 0.9276\ w_1)$ | 95.78 | $s/.975350$ |
| 12 | $.306894\ w_o$ | 81.36 | $.18203\ (w_o + 1.0017\ w_1)$ | 94.97 | $.21177\ (w_o + 0.9231\ w_2)$ | 95.17 | $s/.977559$ |
| 13 | $.299762\ w_o$ | 79.68 | $.17150\ (w_o + 1.0717\ w_1)$ | 94.12 | $.19848\ (w_o + 1.0015\ w_2)$ | 95.00 | $s/.979406$ |
| 14 | $.293534\ w_o$ | 78.09 | $.16244\ (w_o + 1.1381\ w_1)$ | 93.26 | $.18704\ (w_o + 1.0762\ w_2)$ | 94.77 | $s/.980971$ |
| 15 | $.288033\ w_o$ | 76.57 | $.15457\ (w_o + 1.2011\ w_1)$ | 92.39 | $.17708\ (w_o + 1.1477\ w_2)$ | 94.50 | $s/.982316$ |
| 16 | $.283127\ w_o$ | 75.13 | $.14765\ (w_o + 1.2612\ w_1)$ | 91.52 | $.16834\ (w_o + 1.2161\ w_2)$ | 94.18 | $s/.983484$ |
| 17 | $.278716\ w_o$ | 73.76 | $.14153\ (w_o + 1.3186\ w_1)$ | 90.65 | $.16060\ (w_o + 1.2817\ w_2)$ | 93.82 | $s/.984506$ |
| 18 | $.370257\ w_1$ | 72.98 | $.13606\ (w_o + 1.3736\ w_1)$ | 89.78 | $.15369\ (w_o + 1.3448\ w_2)$ | 93.43 | $s/.985410$ |
| 19 | $.362335\ w_1$ | 72.98 | $.13114\ (w_o + 1.4263\ w_1)$ | 88.92 | $.14750\ (w_o + 1.4055\ w_2)$ | 93.02 | $s/.986214$ |
| 20 | $.355214\ w_1$ | 72.91 | $.12670\ (w_o + 1.4769\ w_1)$ | 88.08 | $.14192\ (w_o + 1.4640\ w_2)$ | 92.59 | $s/.986934$ |
| 21 | $.348768\ w_1$ | 72.77 | $.12266\ (w_o + 1.5256\ w_1)$ | 87.24 | $.13684\ (w_o + 1.5206\ w_2)$ | 92.14 | $s/.987583$ |
| 22 | $.342899\ w_1$ | 72.59 | $.11897\ (w_o + 1.5726\ w_1)$ | 86.42 | $.14637\ (w_o + 1.5298\ w_3)$ | 91.78 | $s/.988170$ |
| 23 | $.337526\ w_1$ | 72.37 | $.11558\ (w_o + 1.6179\ w_1)$ | 85.62 | $.14129\ (w_o + 1.5881\ w_3)$ | 91.61 | $s/.988705$ |
| 24 | $.332584\ w_1$ | 72.11 | $.11246\ (w_o + 1.6617\ w_1)$ | 84.83 | $.13663\ (w_o + 1.6446\ w_3)$ | 91.42 | $s/.989193$ |
| 25 | $.328019\ w_1$ | 71.82 | $.10958\ (w_o + 1.7040\ w_1)$ | 84.05 | $.13233\ (w_o + 1.6996\ w_3)$ | 91.21 | $s/.989640$ |

| Sample size, n | Based on one quasi-range | | Based on a linear combination of two adjacent quasi-ranges | | Based on a linear combination of two quasi-ranges among those with $r < r' \le 8$ | | Efficient Estimate |
|---|---|---|---|---|---|---|---|
| | Estimate | Eff(%) | Estimate | Eff(%) | Estimate | Eff(%) | |
| 26 | $.323785\ w_1$ | 71.52 | $.10691\ (w_o + 1.7451\ w_1)$ | 83.29 | $.12836\ (w_o + 1.7529\ w_3)$ | 90.98 | $s/.990052$ |
| 27 | $.319844\ w_1$ | 71.20 | $.10442\ (w_o + 1.7849\ w_1)$ | 82.54 | $.12468\ (w_o + 1.8050\ w_3)$ | 90.73 | $s/.990433$ |
| 28 | $.316165\ w_1$ | 70.86 | $.10209\ (w_o + 1.8235\ w_1)$ | 81.81 | $.12126\ (w_o + 1.8556\ w_3)$ | 90.48 | $s/.990786$ |
| 29 | $.312719\ w_1$ | 70.51 | $.09992\ (w_o + 1.8610\ w_1)$ | 81.10 | $.11807\ (w_o + 1.9050\ w_3)$ | 90.21 | $s/.991113$ |
| 30 | $.309483\ w_1$ | 70.15 | $.09788\ (w_o + 1.8975\ w_1)$ | 80.40 | $.11508\ (w_o + 1.9532\ w_3)$ | 89.93 | $s/.991418$ |
| 31 | $.306436\ w_1$ | 69.78 | $.09596\ (w_o + 1.9329\ w_1)$ | 79.71 | $.11229\ (w_o + 2.0002\ w_3)$ | 89.63 | $s/.991703$ |
| 32 | $.357181\ w_2$ | 69.57 | $.09416\ (w_o + 1.9675\ w_1)$ | 79.04 | $.10967\ (w_o + 2.0461\ w_3)$ | 89.35 | $s/.991969$ |
| 33 | $.353016\ w_2$ | 69.58 | $.09245\ (w_o + 2.0011\ w_1)$ | 78.38 | $.11551\ (w_o + 2.0673\ w_4)$ | 89.11 | $s/.992219$ |
| 34 | $.349094\ w_2$ | 69.57 | $.14484\ (w_1 + 1.2398\ w_2)$ | 78.08 | $.11282\ (w_o + 2.1150\ w_4)$ | 88.97 | $s/.992454$ |
| 35 | $.345394\ w_2$ | 69.53 | $.14201\ (w_1 + 1.2646\ w_2)$ | 77.82 | $.11028\ (w_o + 2.1616\ w_4)$ | 88.82 | $s/.992675$ |
| 36 | $.341895\ w_2$ | 69.48 | $.13934\ (w_1 + 1.2888\ w_2)$ | 77.55 | $.10788\ (w_o + 2.2073\ w_4)$ | 88.66 | $s/.992884$ |
| 37 | $.338580\ w_2$ | 69.41 | $.13680\ (w_1 + 1.3126\ w_2)$ | 77.27 | $.10561\ (w_o + 2.2520\ w_4)$ | 88.48 | $s/.993080$ |
| 38 | $.335433\ w_2$ | 69.32 | $.13440\ (w_1 + 1.3358\ w_2)$ | 76.99 | $.10345\ (w_o + 2.2962\ w_4)$ | 88.31 | $s/.993267$ |
| 39 | $.332442\ w_2$ | 69.21 | $.13211\ (w_1 + 1.3586\ w_2)$ | 76.71 | $.10141\ (w_o + 2.3393\ w_4)$ | 88.12 | $s/.993443$ |
| 40 | $.329593\ w_2$ | 69.10 | $.12995\ (w_1 + 1.3806\ w_2)$ | 76.42 | $.09947\ (w_o + 2.3816\ w_4)$ | 87.92 | $s/.993611$ |
| 41 | $.326875\ w_2$ | 68.97 | $.12788\ (w_1 + 1.4026\ w_2)$ | 76.12 | $.09762\ (w_o + 2.4232\ w_4)$ | 87.73 | $s/.993770$ |
| 42 | $.324280\ w_2$ | 68.83 | $.12592\ (w_1 + 1.4236\ w_2)$ | 75.82 | $.09586\ (w_o + 2.4640\ w_4)$ | 87.52 | $s/.993922$ |
| 43 | $.321798\ w_2$ | 68.68 | $.12403\ (w_1 + 1.4448\ w_2)$ | 75.53 | $.09418\ (w_o + 2.5042\ w_4)$ | 87.32 | $s/.994066$ |
| 44 | $.319420\ w_2$ | 68.53 | $.12223\ (w_1 + 1.4652\ w_2)$ | 75.23 | $.09258\ (w_o + 2.5435\ w_4)$ | 87.10 | $s/.994203$ |
| 45 | $.317141\ w_2$ | 68.37 | $.12051\ (w_1 + 1.4853\ w_2)$ | 74.92 | $.09647\ (w_o + 2.5741\ w_5)$ | 86.97 | $s/.994335$ |
| 46 | $.352208\ w_3$ | 68.22 | $.11886\ (w_1 + 1.5051\ w_2)$ | 74.62 | $.09482\ (w_o + 2.6150\ w_5)$ | 86.85 | $s/.994460$ |
| 47 | $.349387\ w_3$ | 68.24 | $.11727\ (w_1 + 1.5246\ w_2)$ | 74.32 | $.09323\ (w_o + 2.6553\ w_5)$ | 86.73 | $s/.994580$ |
| 48 | $.346682\ w_3$ | 68.23 | $.15270\ (w_2 + 1.1550\ w_3)$ | 74.15 | $.09171\ (w_o + 2.6951\ w_5)$ | 86.60 | $s/.994695$ |
| 49 | $.344086\ w_3$ | 68.22 | $.15056\ (w_2 + 1.1714\ w_3)$ | 74.03 | $.09025\ (w_o + 2.7341\ w_5)$ | 86.47 | $s/.994806$ |
| 50 | $.341592\ w_3$ | 68.20 | $.14850\ (w_2 + 1.1877\ w_3)$ | 73.90 | $.08885\ (w_o + 2.7726\ w_5)$ | 86.33 | $s/.994911$ |

Table 4 (continued)

Most Efficient Unbiased Estimates of Standard Deviation of Normal Population

| Sample size, n | Based on one quasi-range | | Based on a linear combination of two adjacent quasi-ranges | | Based on a linear combination of two quasi-ranges among those with $r < r' \le 8$ | | Efficient Estimate |
|---|---|---|---|---|---|---|---|
| | Estimate | Eff(%) | Estimate | Eff(%) | Estimate | Eff(%) | |
| 51 | .339192 $w_3$ | 68.17 | .14651 ($w_2$ +1.2037 $w_3$) | 73.77 | .08751 ($w_0$ +2.8105 $w_5$) | 86.19 | s/.995013 |
| 52 | .336882 $w_3$ | 68.12 | .14460 ($w_2$ +1.2195 $w_3$) | 73.63 | .08621 ($w_0$ +2.8479 $w_5$) | 86.04 | s/.995110 |
| 53 | .334656 $w_3$ | 68.07 | .14278 ($w_2$ +1.2348 $w_3$) | 73.48 | .08497 ($w_0$ +2.8847 $w_5$) | 85.89 | s/.995204 |
| 54 | .332509 $w_3$ | 68.02 | .14100 ($w_2$ +1.2504 $w_3$) | 73.33 | .08377 ($w_0$ +2.9212 $w_5$) | 85.73 | s/.995294 |
| 55 | .339436 $w_3$ | 67.95 | .13931 ($w_2$ +1.2652 $w_3$) | 73.18 | .08262 ($w_0$ +2.9567 $w_5$) | 85.57 | s/.995381 |
| 56 | .328434 $w_3$ | 67.87 | .13763 ($w_2$ +1.2805 $w_3$) | 73.02 | .13787 ($w_1$ +1.6820 $w_8$) | 85.44 | s/.995465 |
| 57 | .326498 $w_3$ | 67.80 | .13606 ($w_2$ +1.2949 $w_3$) | 72.86 | .13592 ($w_1$ + 1.7062 $w_8$) | 85.46 | s/.995546 |
| 58 | .324625 $w_3$ | 67.71 | .13453 ($w_2$ +1.3093 $w_3$) | 72.70 | .13403 ($w_1$ +1.7303 $w_8$) | 85.47 | s/.995624 |
| 59 | .322812 $w_3$ | 67.62 | .13304 ($w_2$ +1.3236 $w_3$) | 72.53 | .13222 ($w_1$ +1.7539 $w_8$) | 85.48 | s/.995699 |
| 60 | .321055 $w_3$ | 67.53 | .13161 ($w_2$ +1.3374 $w_3$) | 72.36 | .13046 ($w_1$ +1.7774 $w_8$) | 85.48 | s/.995772 |
| 61 | .347463 $w_4$ | 67.52 | .13022 ($w_2$ +1.3514 $w_3$) | 72.19 | .12875 ($w_1$ +1.8006 $w_8$) | 85.47 | s/.995842 |
| 62 | .345399 $w_4$ | 67.52 | .15710 ($w_3$ +1.1114 $w_4$) | 72.06 | .12710 ($w_1$ +1.8236 $w_8$) | 85.46 | s/.995910 |
| 63 | .343400 $w_4$ | 67.52 | .15534 ($w_3$ +1.1243 $w_4$) | 71.99 | .12551 ($w_1$ +1.8461 $w_8$) | 85.44 | s/.995976 |
| 64 | .341461 $w_4$ | 67.50 | .15370 ($w_3$ +1.1360 $w_4$) | 71.91 | .12397 ($w_1$ +1.8684 $w_8$) | 85.42 | s/.996040 |
| 65 | .339581 $w_4$ | 67.49 | .15210 ($w_3$ +1.1478 $w_4$) | 71.83 | .12247 ($w_1$ +1.8907 $w_8$) | 85.40 | s/.996102 |
| 66 | .337755 $w_4$ | 67.46 | .15052 ($w_3$ +1.1598 $w_4$) | 71.75 | .12102 ($w_1$ +1.9125 $w_8$) | 85.36 | s/.996161 |
| 67 | .335982 $w_4$ | 67.44 | .14895 ($w_3$ +1.1723 $w_4$) | 71.67 | .11961 ($w_1$ +1.9342 $w_8$) | 85.33 | s/.996219 |
| 68 | .334260 $w_4$ | 67.40 | .14750 ($w_3$ +1.1835 $w_4$) | 71.57 | .11826 ($w_1$ +1.9554 $w_8$) | 85.29 | s/.996276 |
| 69 | .332585 $w_4$ | 67.36 | .14607 ($w_3$ +1.1949 $w_4$) | 71.48 | .11693 ($w_1$ +1.9765 $w_8$) | 85.24 | s/.996330 |
| 70 | .330956 $w_4$ | 67.32 | .14466 ($w_3$ +1.2064 $w_4$) | 71.39 | .11564 ($w_1$ +1.9975 $w_8$) | 85.20 | s/.996383 |
| 71 | .329370 $w_4$ | 67.27 | .14333 ($w_3$ +1.2172 $w_4$) | 71.29 | .11439 ($w_1$ +2.0181 $w_8$) | 85.14 | s/.996435 |
| 72 | .327827 $w_4$ | 67.22 | .14201 ($w_3$ +1.2284 $w_4$) | 71.18 | .11317 ($w_1$ +2.0387 $w_8$) | 85.09 | s/.996485 |
| 73 | .326323 $w_4$ | 67.16 | .14071 ($w_3$ +1.2395 $w_4$) | 71.08 | .11198 ($w_1$ +2.0593 $w_8$) | 85.04 | s/.996534 |
| 74 | .324857 $w_4$ | 67.11 | .13943 ($w_3$ +1.2510 $w_4$) | 70.98 | .11083 ($w_1$ +2.0793 $w_8$) | 84.98 | s/.996581 |
| 75 | .346274 $w_5$ | 67.07 | .13821 ($w_3$ +1.2617 $w_4$) | 70.87 | .10971 ($w_1$ +2.0992 $w_8$) | 84.91 | s/.996627 |
| 76 | .344605 $w_5$ | 67.08 | .13703 ($w_3$ +1.2723 $w_4$) | 70.75 | .10862 ($w_1$ +2.1188 $w_8$) | 84.85 | s/.996672 |
| 77 | .342979 $w_5$ | 67.07 | .15847 ($w_4$ +1.0948 $w_5$) | 70.71 | .10756 ($w_1$ +2.1382 $w_8$) | 84.78 | s/.996716 |
| 78 | .341393 $w_5$ | 67.07 | .15704 ($w_4$ +1.1049 $w_5$) | 70.66 | .10653 ($w_1$ +2.1575 $w_8$) | 84.71 | s/.996759 |
| 79 | .339847 $w_5$ | 67.06 | .15568 ($w_4$ +1.1145 $w_5$) | 70.61 | .10552 ($w_1$ +2.1766 $w_8$) | 84.63 | s/.996800 |
| 80 | .338338 $w_5$ | 67.04 | .15431 ($w_4$ +1.1245 $w_5$) | 70.56 | .10453 ($w_1$ +2.1956 $w_8$) | 84.56 | s/.996841 |
| 81 | .336865 $w_5$ | 67.03 | .15305 ($w_4$ +1.1334 $w_5$) | 70.50 | .10357 ($w_1$ + 2.2145 $w_8$) | 84.48 | s/.996880 |
| 82 | .335427 $w_5$ | 67.01 | .15176 ($w_4$ +1.1432 $w_5$) | 70.44 | .10264 ($w_1$ + 2.2328 $w_8$) | 84.40 | s/.996918 |
| 83 | .334022 $w_5$ | 66.98 | .15056 ($w_4$ +1.1519 $w_5$) | 70.38 | .10171 ($w_1$ + 2.2515 $w_8$) | 84.32 | s/.996956 |
| 84 | .332649 $w_5$ | 66.95 | .14934 ($w_4$ +1.1612 $w_5$) | 70.31 | .10082 ($w_1$ + 2.2696 $w_8$) | 84.23 | s/.996993 |
| 85 | .331306 $w_5$ | 66.92 | .14809 ($w_4$ +1.1714 $w_5$) | 70.25 | .09996 ($w_1$ + 2.2875 $w_8$) | 84.15 | s/.997028 |
| 86 | .329994 $w_5$ | 66.89 | .14697 ($w_4$ +1.1800 $w_5$) | 70.18 | .09910 ($w_1$ +2.3057 $w_8$) | 84.07 | s/.997063 |
| 87 | .328710 $w_5$ | 66.85 | .14584 ($w_4$ +1.1890 $w_5$) | 70.11 | .09826 ($w_1$ +2.3234 $w_8$) | 83.97 | s/.997097 |
| 88 | .327454 $w_5$ | 66.81 | .14475 ($w_4$ +1.1977 $w_5$) | 70.03 | .09746 ($w_1$ +2.3406 $w_8$) | 83.88 | s/.997131 |
| 89 | .345465 $w_6$ | 66.77 | .14363 ($w_4$ +1.2071 $w_5$) | 69.96 | .09665 ($w_1$ +2.3585 $w_8$) | 83.79 | s/.997163 |
| 90 | .344065 $w_6$ | 66.77 | .14259 ($w_4$ +1.2156 $w_5$) | 69.88 | .09588 ($w_1$ +2.3755 $w_8$) | 83.70 | s/.997195 |
| 91 | .342694 $w_6$ | 66.77 | .16064 ($w_5$ +1.0751 $w_6$) | 69.82 | .09511 ($w_1$ +2.3930 $w_8$) | 83.61 | s/.997226 |
| 92 | .341353 $w_6$ | 66.77 | .15934 ($w_5$ +1.0844 $w_6$) | 69.79 | .09437 ($w_1$ +2.4096 $w_8$) | 83.51 | s/.997257 |
| 93 | .340040 $w_6$ | 66.76 | .15823 ($w_5$ +1.0916 $w_6$) | 69.75 | .09364 ($w_1$ +2.4264 $w_8$) | 83.42 | s/.997286 |
| 94 | .338754 $w_6$ | 66.75 | .15705 ($w_5$ +1.0999 $w_6$) | 69.71 | .09293 ($w_1$ +2.4431 $w_8$) | 83.32 | s/.997315 |
| 95 | .337494 $w_6$ | 66.74 | .15592 ($w_5$ +1.1078 $w_6$) | 69.67 | .09223 ($w_1$ +2.4595 $w_8$) | 83.22 | s/.997344 |
| 96 | .336259 $w_6$ | 66.73 | .15480 ($w_5$ +1.1158 $w_6$) | 69.63 | .09155 ($w_1$ +2.4761 $w_8$) | 83.12 | s/.997371 |
| 97 | .335049 $w_6$ | 66.71 | .15366 ($w_5$ +1.1244 $w_6$) | 69.59 | .09087 ($w_1$ +2.4924 $w_8$) | 83.02 | s/.997399 |
| 98 | .333863 $w_6$ | 66.69 | .15261 ($w_5$ +1.1319 $w_6$) | 69.54 | .09022 ($w_1$ +2.5085 $w_8$) | 82.92 | s/.997426 |
| 99 | .332700 $w_6$ | 66.67 | .15161 ($w_5$ +1.1389 $w_6$) | 69.49 | .08957 ($w_1$ +2.5245 $w_8$) | 82.82 | s/.997452 |
| 100 | .331559 $w_6$ | 66.65 | .15059 ($w_5$ +1.1466 $w_6$) | 69.44 | .08895 ($w_1$ +2.5401 $w_8$) | 82.71 | s/.997478 |

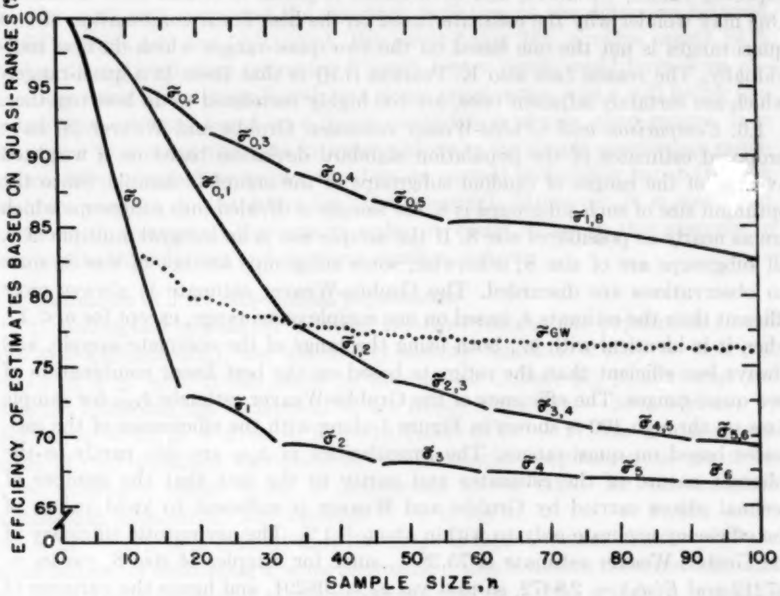


**Fig. 1. Efficiency of Estimates of Standard Deviation for Normal Population**

most efficient estimate based on one sample quasi-range, together with its efficiency, for $n = 2\,(1)\,100$, also the most efficient estimates based on linear combinations of two adjacent quasi-ranges and of two quasi-ranges among those with $r < r' \leqq 8$, together with their efficiencies, for $n = 4\,(1)\,100$, and the efficient estimate based on the sample standard deviation. For the estimates based on one sample quasi-range, the numerical coefficients $1/E(w_r)$ are accurate to within a unit in the sixth decimal place, and the efficiencies are accurate to within 0.01 %. For the estimates based on a linear combination of two sample quasi-ranges, the numerical coefficients $1/E(w_r + \lambda_{r,r'}w_{r'})$ are accurate to within a unit in the fourth decimal place, the values of $\lambda_{r,r'}$ are accurate to within a unit in the third decimal place, and the efficiencies are accurate to within 0.01 %. The efficiency of the estimates based on quasi-ranges is shown graphically by Figure 1. It will be noted that for $n < 56$ the estimate based on the best linear combination of two quasi-ranges among those with $r < r' \leqq 8$ always involves the range ($r = 0$), with $1 \leqq r' \leqq 5$ while the best such estimate for $56 \leqq n \leqq 100$ is $\tilde\sigma_{1,8}$, with the efficiency dropping to 82.71 % for $n = 100$. It seems likely that slightly better estimates for $n$ near 100 could be obtained by dropping the restriction $r' \leqq 8$, but it is doubtful whether the increase in efficiency would exceed 1 %, which would hardly justify the additional computation of

expected values, variances, and covariances required to obtain such estimates. One may wonder why the estimate based on the best linear combination of two quasi-ranges is not the one based on the two quasi-ranges which do best individually. The reason (see also K. Pearson [15]) is that these two quasi-ranges, which are certainly adjacent ones, are too highly correlated to do best together.

1.6. *Comparison with Grubbs-Weaver estimates.* Grubbs and Weaver [8] have proposed estimates of the population standard deviation based on a weighted average of the ranges of random subgroups of the complete sample. Since the optimum size of such subgroups is 8, the sample is divided into subgroups which are as nearly as possible of size 8. If the sample size is an integral multiple of 8, all subgroups are of size 8; otherwise, some subgroups are not of size 8, since no observations are discarded. The Grubbs-Weaver estimate is always more efficient than the estimate $\bar{\sigma}_r$ based on one sample quasi-range, except for $n < 12$, when it is identical with $\bar{\sigma}_r$, both using the range of the complete sample, and always less efficient than the estimate based on the best linear combination of two quasi-ranges. The efficiency of the Grubbs-Weaver estimate $\bar{\sigma}_{GW}$ for sample sizes up through 100 is shown in Figure 1 along with the efficiencies of the estimates based on quasi-ranges. The irregularities in $\bar{\sigma}_{GW}$ are due partly to the inherent nature of the estimates and partly to the fact that the number of decimal places carried by Grubbs and Weaver is sufficient to yield values of the efficiency accurate only to within about 0.1%. The asymptotic efficiency of the Grubbs-Weaver estimate is 75.38%, since for samples of size 8, var $w_0$ = .67212 and $E(w_0) = 2.8472$, so that var $\bar{\sigma}_0 = .08291$, and hence the variance of $\bar{\sigma}_{GW}$, which is the mean of $n/8$ such estimates, is $.08291/(n/8) = .6633/n$, as compared with an asymptotic variance $1/2n = .5/n$ for $\hat{\sigma}$. By using results given by K. Pearson [15] and by Benson [1], one can easily show that the corresponding asymptotic efficiencies are 65.23% for estimates based on one quasi-range and also for those based on the best linear combination of two adjacent quasi-ranges, and approximately 80.08% for estimates based on the best linear combination of two quasi-ranges.

1.7. *Example.* As an example of the use of estimates based on sample quasi-ranges, consider the following data, given by Morse and Kimball ([10], p. 134) and assumed to come from a normal population, which represent the deviation (in one dimension) from the aiming point of the mean point of impact of salvos of two projectiles:

| | | |
|---|---|---|
| −237 | −23 | Quasi-ranges: |
| −133 | −13 | $w_0 = 270 - (-237) = 507$ |
| −93 | −10 | $w_1 = 209 - (-133) = 342$ |
| −77 | 57 | $w_2 = 173 - (-93) = 266$ |
| −75 | 65 | Sample standard deviation: $s = 127.2$ |
| −70 | 142 | Estimates of population standard deviation: |
| −66 | 154 | $\bar{\sigma}_1 = .355214\, w_1 = 121.5$ |
| −65 | 173 | $\bar{\sigma}_{0,1} = .12670\, (w_0 + 1.4769\, w_1) = 128.2$ |
| −34 | 209 | $\bar{\sigma}_{0,2} = .14192\, (w_0 + 1.4640\, w_1) = 127.2$ |
| −28 | 270 | $\hat{\sigma} = s/.986934 = 128.9$ |

Morse and Kimball plotted the data on normal probability paper, fitted a straight line "by eye", and estimated the standard deviation as the difference between the 84% point and the 50% point, the result being 161, a value nearly 25% greater than the efficient estimate. A much better result could have been obtained by using an estimate based on a single quasi-range, and a still better one by using an estimate based on a linear combination of two quasi-ranges. It is also easier to arrange the data in order and make the simple quasi-range calculations shown above than to plot the data on normal probability paper (though one may want to do the latter for other reasons). Moreover, it is really not necessary to arrange all the data in order; it would suffice in this example to pick out the three largest and the three smallest values.

**2. Estimates of $\sigma$ for non-normal populations.**

2.1. *Rectangular population.* For the standard rectangular population (mean zero and variance one), the probability density function is $f(x) = 1/2\sqrt{3}$, $-\sqrt{3} \leqq x \leqq \sqrt{3}$. It can easily be shown (see Cramér [3], p. 372) that the expected value and the variance of $w_r$ are $E(w_r) = 2\sqrt{3}(n - 2r - 1)/(n + 1)$ and $\text{var}\, w_r = 12(2r + 2)(n - 2r - 1)/(n + 1)^2(n + 2)$. An unbiased estimate of the standard deviation of a rectangular population is given by $\tilde{\sigma}_r = w_r/E(w_r)$, where $E(w_r)$ is understood to be taken for the standard rectangular population. The variance of $\tilde{\sigma}_r$ is $\text{var}\, \tilde{\sigma}_r = (2r + 2)/[(n + 2)(n - 2r - 1)]$. It is evident that the range is more efficient than any of the quasi-ranges for estimating $\sigma$, since increasing $r$ both increases the numerator and decreases the denominator of the expression for $\text{var}\, \tilde{\sigma}_r$. As a matter of fact, it can be shown that the range is an efficient statistic for estimating the standard deviation of a rectangular population. Table 5 gives the unbiased estimates $\tilde{\sigma}_0$ for $n = 2\ (1)\ 100$. The numerical coefficients $1/E(w_0)$ are accurate to within a unit in the sixth decimal place. Since the efficiency is always 100%, it is not given in the table.

2.2. *Exponential population.* For the exponential population with mean and variance each equal to one, the probability density function is $f(x) = e^{-x}$, $0 \leqq x < \infty$. Rider [16] has shown that the expected value and the variance of $w_r$ for samples of $n$ from this population are

$$E(w_r) = \sum_{j=r+1}^{n-r-1} \frac{1}{j} \tag{3}$$

and

$$\text{var}\, w_r = \sum_{j=r+1}^{n-r-1} \frac{1}{j^2}. \tag{4}$$

An unbiased estimate of $\sigma$ for an exponential population is $\tilde{\sigma}_r = w_r/E(w_r)$ where $E(w_r)$ is understood to be taken for an exponential population with variance one. The variance of $\tilde{\sigma}_r$ is

$$\text{var}\, \tilde{\sigma}_r = \sum_{j=r+1}^{n-r-1} \frac{1}{j^2} \Big/ \left(\sum_{j=r+1}^{n-r-1} \frac{1}{j}\right)^2. \tag{5}$$

TABLE 5

ESTIMATES OF STANDARD DEVIATION OF RECTANGULAR POPULATION

| Sample size, n | Estimate based on the range | Sample size, n | Estimate based on the range | Sample size, n | Estimate based on the range |
|---|---|---|---|---|---|
| | | 36 | .305171 $w_o$ | 71 | .296923 $w_o$ |
| 2 | .866025 $w_o$ | 37 | .304713 $w_o$ | 72 | .296807 $w_o$ |
| 3 | .577350 $w_o$ | 38 | .304279 $w_o$ | 73 | .296694 $w_o$ |
| 4 | .481125 $w_o$ | 39 | .303869 $w_o$ | 74 | .296584 $w_o$ |
| 5 | .433013 $w_o$ | 40 | .303479 $w_o$ | 75 | .296477 $w_o$ |
| 6 | .404145 $w_o$ | 41 | .303109 $w_o$ | 76 | .296373 $w_o$ |
| 7 | .384900 $w_o$ | 42 | .302757 $w_o$ | 77 | .296272 $w_o$ |
| 8 | .371154 $w_o$ | 43 | .302422 $w_o$ | 78 | .296173 $w_o$ |
| 9 | .360844 $w_o$ | 44 | .302102 $w_o$ | 79 | .296077 $w_o$ |
| 10 | .352825 $w_o$ | 45 | .301797 $w_o$ | 80 | .295983 $w_o$ |
| 11 | .346410 $w_o$ | 46 | .301505 $w_o$ | 81 | .295892 $w_o$ |
| 12 | .341162 $w_o$ | 47 | .301226 $w_o$ | 82 | .295803 $w_o$ |
| 13 | .336788 $w_o$ | 48 | .300959 $w_o$ | 83 | .295716 $w_o$ |
| 14 | .333087 $w_o$ | 49 | .300703 $w_o$ | 84 | .295631 $w_o$ |
| 15 | .329914 $w_o$ | 50 | .300458 $w_o$ | 85 | .295548 $w_o$ |
| 16 | .327165 $w_o$ | 51 | .300222 $w_o$ | 86 | .295467 $w_o$ |
| 17 | .324760 $w_o$ | 52 | .299996 $w_o$ | 87 | .295389 $w_o$ |
| 18 | .322637 $w_o$ | 53 | .299778 $w_o$ | 88 | .295311 $w_o$ |
| 19 | .320750 $w_o$ | 54 | .299569 $w_o$ | 89 | .295236 $w_o$ |
| 20 | .319062 $w_o$ | 55 | .299367 $w_o$ | 90 | .295162 $w_o$ |
| 21 | .317543 $w_o$ | 56 | .299172 $w_o$ | 91 | .295090 $w_o$ |
| 22 | .316168 $w_o$ | 57 | .298985 $w_o$ | 92 | .295020 $w_o$ |
| 23 | .314918 $w_o$ | 58 | .298804 $w_o$ | 93 | .294951 $w_o$ |
| 24 | .313777 $w_o$ | 59 | .298629 $w_o$ | 94 | .294883 $w_o$ |
| 25 | .312731 $w_o$ | 60 | .298461 $w_o$ | 95 | .294817 $w_o$ |
| 26 | .311769 $w_o$ | 61 | .298298 $w_o$ | 96 | .294753 $w_o$ |
| 27 | .310881 $w_o$ | 62 | .298140 $w_o$ | 97 | .294689 $w_o$ |
| 28 | .310058 $w_o$ | 63 | .297987 $w_o$ | 98 | .294627 $w_o$ |
| 29 | .309295 $w_o$ | 64 | .297839 $w_o$ | 99 | .294566 $w_o$ |
| 30 | .308584 $w_o$ | 65 | .297696 $w_o$ | 100 | .294507 $w_o$ |
| 31 | .307920 $w_o$ | 66 | .297557 $w_o$ | | |
| 32 | .307299 $w_o$ | 67 | .297423 $w_o$ | | |
| 33 | .306717 $w_o$ | 68 | .297292 $w_o$ | | |
| 34 | .306171 $w_o$ | 69 | .297166 $w_o$ | | |
| 35 | .305656 $w_o$ | 70 | .297043 $w_o$ | | |

For the exponential population with mean and standard deviation each equal to $c$, the probability density function is $f(x) = (1/c)e^{-x/c}$, $0 \leq x < \infty$. The sample mean $\bar{x}$ is the efficient estimate of the parameter $c$, and has variance $c^2/n$. When $c = 1$, var $\bar{x} = 1/n$. Thus the efficiency, for an exponential population whose lower limit is zero (or some other known value $x_0$), of the estimate $\bar{\sigma}_r$ based on the $r$th quasi-range is given by the ratio of the variance of the efficient estimate $\bar{x}$(or $\bar{x} - x_0$) to the variance of $\bar{\sigma}_r$, that is by

$$\text{Eff } \tilde{\sigma}_r = \left(\sum_{j=r+1}^{n-r-1} \frac{1}{j}\right)^2 \Big/ n \sum_{j=r+1}^{n-r-1} \frac{1}{j^2}. \tag{6}$$

The most efficient estimates $\tilde{\sigma}_r$, together with their efficiencies, have been computed for $n = 2\ (1)\ 100$, but they will not be tabulated here, since the efficiencies are somewhat disappointing, varying from 50.00% to 61.73%. It should not be surprising that quasi-ranges, which are differences of symmetric order statistics, are not very efficient in estimating the standard deviation of an unsymmetric population. It is interesting, however, to note that the standard deviation of an exponential population whose lower limit is known can be estimated more efficiently from a single order statistic. The author is currently investigating the efficiency of estimates based on a linear combination of two order statistics, and preliminary results look promising,

2.3. *Bias when estimates which assume normality are used.* Paragraphs 2.1 and 2.2 cover the cases in which the population being sampled is known to be rectangular or exponential. Suppose, however, that the population is of one or the other of these two types, but the investigator who is interested in estimating the standard deviation is not aware of this fact, and proceeds to use one of the estimates which assume normality. In this case, the estimate is no longer unbiased. The bias of an estimate, based on one sample quasi-range, which assumes normality, when the population being sampled is actually of some other type, is given by

$$B_0 = [E_0(w_r) - E_n(w_r)]/E_n(w_r). \tag{7}$$

The bias of an estimate, based on a linear combination of two sample quasi-ranges, which assumes normality, when the population being sampled is actually of some other type, is given by

$$B_0 = \frac{[E_0(w_r) + \lambda_{r,r'} E_0(w_{r'})] - [E_n(w_r) + \lambda_{r,r'} E_n(w_{r'})]}{E_n(w_r) + \lambda_{r,r'} E_n(w_{r'})}. \tag{8}$$

In equations (7) and (8), $E_n$ represents an expected value taken for the normal population, while $E_0$ represents an expected value taken for the other population, both populations having variance one. Table 6 gives the bias $B_r$ for a rectangular population and the bias $B_e$ for an exponential population when the estimates of Table 4, which assume normality, are used. In both cases, the values of the bias are accurate to within 0.01%.

*Acknowledgements.* The author gratefully acknowledges the assistance rendered by Dr. Paul R. Rider, Dr. Gertrude Blanch, the Editor, and the referee, all of whom made helpful suggestions, and by Eugene H. Guthrie, who did the major share of the programming of computations for the Univac Scientific (ERA 1103) computer and monitored the work of the computer operators.

TABLE 6

Bias (%) of Estimates which Assume Normality

| | When Population is Rectangular | | | When Population is Exponential | | |
|---|---|---|---|---|---|---|
| Sample size, n | One quasi-range | Two adjacent quasi-ranges | Two quasi-ranges with $r < r' \le 8$ | One quasi-range | Two adjacent quasi-ranges | Two quasi-ranges with $r < r' \le 8$ |
| 2 | 2.33 | | | -11.38 | | |
| 3 | 2.33 | | | -11.38 | | |
| 4 | 0.96 | 1.98 | 1.98 | -10.95 | -11.27 | -11.27 |
| 5 | -0.71 | 1.61 | 1.61 | -10.43 | -11.15 | -11.16 |
| 6 | -2.37 | 1.13 | 1.13 | -9.91 | -11.01 | -11.01 |
| 7 | -3.93 | 0.54 | 0.54 | -9.41 | -10.84 | -10.84 |
| 8 | -5.37 | -0.11 | -0.11 | -8.93 | -10.66 | -10.66 |
| 9 | -6.69 | -0.80 | -0.80 | -8.49 | -10.46 | -10.46 |
| 10 | -7.90 | -1.51 | -1.51 | -8.08 | -10.26 | -10.26 |
| 11 | -9.02 | -2.22 | -2.22 | -7.69 | -10.06 | -10.06 |
| 12 | -10.04 | -2.92 | -1.47 | -7.32 | -9.85 | -10.07 |
| 13 | -10.99 | -3.60 | -1.72 | -6.98 | -9.66 | -10.00 |
| 14 | -11.87 | -4.27 | -2.01 | -6.65 | -9.46 | -9.92 |
| 15 | -12.69 | -4.92 | -2.33 | -6.34 | -9.27 | -9.84 |
| 16 | -13.46 | -5.54 | -2.66 | -6.05 | -9.08 | -9.75 |
| 17 | -14.18 | -6.14 | -3.01 | -5.77 | -8.89 | -9.65 |
| 18 | 1.26 | -6.72 | -3.37 | -11.85 | -8.71 | -9.56 |
| 19 | 0.41 | -7.28 | -3.74 | -11.61 | -8.53 | -9.46 |
| 20 | -0.39 | -7.82 | -4.11 | -11.37 | -8.36 | -9.36 |
| 21 | -1.15 | -8.33 | -4.48 | -11.14 | -8.20 | -9.25 |
| 22 | -1.87 | -8.83 | -3.11 | -10.92 | -8.03 | -9.43 |
| 23 | -2.56 | -9.31 | -3.32 | -10.71 | -7.87 | -9.38 |
| 24 | -3.22 | -9.78 | -3.53 | -10.51 | -7.72 | -9.33 |
| 25 | -3.85 | -10.23 | -3.75 | -10.31 | -7.56 | -9.28 |
| 26 | -4.45 | -10.66 | -3.98 | -10.12 | -7.42 | -9.22 |
| 27 | -5.03 | -11.08 | -4.21 | -9.93 | -7.27 | -9.16 |
| 28 | -5.58 | -11.48 | -4.45 | -9.75 | -7.13 | -9.10 |
| 29 | -6.11 | -11.87 | -4.69 | -9.58 | -6.99 | -9.04 |
| 30 | -6.63 | -12.25 | -4.94 | -9.41 | -6.86 | -8.98 |
| 31 | -7.12 | -12.61 | -5.18 | -9.24 | -6.73 | -8.91 |
| 32 | 1.23 | -12.97 | -5.43 | -12.07 | -6.60 | -8.85 |
| 33 | .0.71 | -13.31 | -3.95 | -11.92 | -6.47 | -9.09 |
| 34 | 0.20 | -4.02 | -4.11 | -11.78 | -10.32 | -9.05 |
| 35 | -0.29 | -4.43 | -4.28 | -11.63 | -10.19 | -9.02 |

TABLE 6
(Continued)
Bias (%) of Estimates which Assume Normality

| | When Population is Rectangular | | | When Population is Exponential | | |
|---|---|---|---|---|---|---|
| Sample size, n | One quasi-range | Two adjacent quasi-ranges | Two quasi-ranges with r < r'≤8 | One quasi-ranges | Two adjacent quasi-ranges | Two quasi-ranges with r<r'≤ 8 |
| 36 | -0.77 | -4.83 | -4.45 | -11.49 | -10.06 | -8.98 |
| 37 | -1.23 | -5.22 | -4.63 | -11.35 | -9.94 | -8.93 |
| 38 | -1.68 | -5.60 | -4.81 | -11.22 | -9.82 | -8.89 |
| 39 | -2.11 | -5.96 | -4.99 | -11.09 | -9.70 | -8.85 |
| 40 | -2.53 | -6.32 | -5.17 | -10.96 | -9.58 | -8.81 |
| 41 | -2.94 | -6.67 | -5.36 | -10.83 | -9.47 | -8.76 |
| 42 | -3.34 | -7.00 | -5.54 | -10.71 | -9.36 | -8.72 |
| 43 | -3.73 | -7.33 | -5.72 | -10.59 | -9.25 | -8.67 |
| 44 | -4.10 | -7.65 | -5.91 | -10.47 | -9.14 | -8.62 |
| 45 | -4.47 | -7.97 | -4.45 | -10.35 | -9.03 | -8.89 |
| 46 | 1.24 | -8.27 | -4.59 | -12.18 | -8.93 | -8.86 |
| 47 | 0.86 | -8.57 | -4.73 | -12.07 | -8.83 | -8.83 |
| 48 | 0.49 | -2.46 | -4.88 | -11.96 | -11.01 | -8.80 |
| 49 | 0.12 | -2.79 | -5.02 | -11.86 | -10.91 | -8.76 |
| 50 | -0.23 | -3.10 | -5.17 | -11.75 | -10.81 | -8.73 |
| 51 | -0.58 | -3.41 | -5.32 | -11.65 | -10.72 | -8.70 |
| 52 | -0.92 | -3.71 | -5.46 | -11.55 | -10.62 | -8.66 |
| 53 | -1.25 | -4.01 | -5.61 | -11.45 | -10.53 | -8.63 |
| 54 | -1.57 | -4.30 | -5.76 | -11.36 | -10.44 | -8.59 |
| 55 | -1.89 | -4.58 | -5.91 | -11.26 | -10.35 | -8.55 |
| 56 | -2.20 | -4.86 | -0.63 | -11.17 | -10.26 | -10.80 |
| 57 | -2.50 | -5.13 | -0.76 | -11.08 | -10.17 | -10.78 |
| 58 | -2.79 | -5.40 | -0.89 | -10.98 | -10.09 | -10.75 |
| 59 | -3.08 | -5.66 | -1.02 | -10.89 | -10.00 | -10.71 |
| 60 | -3.37 | -5.91 | -1.15 | -10.81 | -9.92 | -10.67 |
| 61 | 0.95 | -6.16 | -1.28 | -12.16 | -9.83 | -10.64 |
| 62 | 0.66 | -1.61 | -1.41 | -12.07 | -11.36 | -10.60 |
| 63 | 0.37 | -1.87 | -1.55 | -11.99 | -11.28 | -10.56 |
| 64 | 0.09 | -2.13 | -1.67 | -11.91 | -11.20 | -10.53 |
| 65 | -0.19 | -2.39 | -1.81 | -11.83 | -11.13 | -10.49 |
| 66 | -0.46 | -2.64 | -1.94 | -11.75 | -11.05 | -10.46 |
| 67 | -0.73 | -2.88 | -2.07 | -11.67 | -10.97 | -10.42 |
| 68 | -0.99 | -3.12 | -2.20 | -11.59 | -10.90 | -10.38 |
| 69 | -1.25 | -3.36 | -2.34 | -11.51 | -10.83 | -10.35 |
| 70 | -1.50 | -3.59 | -2.47 | -11.44 | -10.75 | -10.31 |

TABLE 6
(Continued)
Bias ( % ) of Estimates which Assume Normality

| Sample size, n | When Population is Rectangular: One quasi-range | When Population is Rectangular: Two adjacent quasi-ranges | When Population is Rectangular: Two quasi-ranges with $r < r' \leq 8$ | When Population is Exponential: One quasi-ranges | When Population is Exponential: Two adjacent quasi-ranges | When Population is Exponential: Two quasi-ranges with $r < r' \leq 8$ |
|---|---|---|---|---|---|---|
| 71 | -1.75 | -3.82 | -2.59 | -11.36 | -10.68 | -10.27 |
| 72 | -1.99 | -4.05 | -2.73 | -11.29 | -10.61 | -10.24 |
| 73 | -2.23 | -4.27 | -2.85 | -11.22 | -10.54 | -10.20 |
| 74 | -2.47 | -4.49 | -2.98 | -11.14 | -10.47 | -10.17 |
| 75 | 1.01 | -4.70 | -3.11 | -12.21 | -10.40 | -10.13 |
| 76 | 0.77 | -4.91 | -3.24 | -12.14 | -10.33 | -10.09 |
| 77 | 0.53 | -1.29 | -3.36 | -12.07 | -11.51 | -10.05 |
| 78 | 0.30 | -1.51 | -3.49 | -12.01 | -11.45 | -10.02 |
| 79 | 0.07 | -1.73 | -3.62 | -11.94 | -11.38 | -9.98 |
| 80 | -0.16 | -1.94 | -3.74 | -11.87 | -11.32 | -9.95 |
| 81 | -0.38 | -2.15 | -3.87 | -11.81 | -11.25 | -9.91 |
| 82 | -0.60 | -2.36 | -3.99 | -11.74 | -11.19 | -9.88 |
| 83 | -0.82 | -2.56 | -4.11 | -11.68 | -11.13 | -9.84 |
| 84 | -1.04 | -2.76 | -4.24 | -11.62 | -11.07 | -9.81 |
| 85 | -1.25 | -2.96 | -4.36 | -11.55 | -11.01 | -9.77 |
| 86 | -1.45 | -3.15 | -4.48 | -11.49 | -10.95 | -9.74 |
| 07 | -1.66 | -3.34 | -4.60 | -11.43 | -10.89 | -9.70 |
| 88 | -1.86 | -3.53 | -4.72 | -11.37 | -10.83 | -9.67 |
| 89 | 1.06 | -3.72 | -4.83 | -12.25 | -10.77 | -9.63 |
| 90 | 0.85 | -3.91 | -4.95 | -12.19 | -10.71 | -9.60 |
| 91 | 0.65 | -0.89 | -5.07 | -12.13 | -11.67 | -9.56 |
| 92 | 0.45 | -1.08 | -5.19 | -12.08 | -11.61 | -9.53 |
| 93 | 0.25 | -1.27 | -5.30 | -12.02 | -11.55 | -9.49 |
| 94 | 0.05 | -1.45 | -5.42 | -11.96 | -11.50 | -9.46 |
| 95 | -0.14 | -1.63 | -5.53 | -11.91 | -11.45 | -9.43 |
| 96 | -0.33 | -1.81 | -5.64 | -11.85 | -11.39 | -9.39 |
| 97 | -0.52 | -1.99 | -5.76 | -11.80 | -11.34 | -9.36 |
| 98 | -0.70 | -2.17 | -5.87 | -11.74 | -11.28 | -9.32 |
| 99 | -0.88 | -2.34 | -5.98 | -11.69 | -11.23 | -9.29 |
| 100 | -1.07 | -2.51 | -6.09 | -11.63 | -11.18 | -9.26 |

## REFERENCES


[1] F. Benson, "A note on the estimation of mean and standard deviation from quantiles," *J. Roy. Stat. Soc.*, Ser. B, Vol. 11 (1949), pp. 91–100.

[2] J. H. Cadwell, "The distribution of quasi-ranges in samples from a normal population," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 603–613.

[3] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[4] W. J. Dixon, "Estimates of the mean and standard deviation of a normal population," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 806–809.

[5] F. Y. Edgeworth, "Methods of statistics," *J. Roy. Stat. Soc.*, Jubilee Volume (1885), pp. 181–217.

[6] H. J. Godwin, "On the estimation of dispersion by linear systematic statistics," *Biometrika*, Vol. 36 (1949), pp. 92–100.

[7] H. J. Godwin, "Some low moments of order statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 279–285.

[8] FRANK E. GRUBBS AND CHALMERS L. WEAVER, "The best unbiased estimate of population standard deviation based on group ranges," *J. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 224–241.

[9] A. E. JONES, "A useful method for the routine estimation of dispersion from large samples," *Biometrika*, Vol. 33 (1946), pp. 274–282.

[10] PHILIP M. MORSE AND GEORGE E. KIMBALL, *Methods of Operations Research*, Technology Press, Cambridge, 1950.

[11] FREDERICK MOSTELLER, "On some useful inefficient statistics," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 377–408.

[12] K. R. NAIR, "A note on the mean deviation from the median," *Biometrika*, Vol. 34 (1947), pp. 360–362.

[13] K. R. NAIR, "Efficiencies of certain linear systematic statistics for estimating dispersion from normal samples," *Biometrika*, Vol. 37 (1950), pp. 182–183.

[14] EGON S. PEARSON, "The percentage limits for the distribution of range in samples from a normal population ($n < 100$)," *Biometrika*, Vol. 24 (1932), pp. 404–417.

[15] KARL PEARSON, "On the probable errors of frequency constants, Part III," *Biometrika*, Vol. 13 (1920), pp. 113–132.

[16] PAUL R. RIDER, "Quasi-ranges of samples from an exponential population," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 252–254.

[17] A. E. SARHAN AND B. G. GREENBERG, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples, Part I," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 427–451.

[18] W. F. SHEPPARD, "On the application of the theory of error to cases of normal distribution and normal correlation," *Philos. Trans. Roy. Soc. London*, Ser. A, Vol. 192 (1899), pp. 101–167.

[19] D. TEICHROEW, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.

[20] L. H. C. TIPPETT "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 17 (1925), pp. 364–387.

[21] S. S. WILKS, "Order statistics," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 6–50.

# THE JOINT CUMULANTS OF TRUE VALUES AND ERRORS OF MEASUREMENT

By Frederic M. Lord

*Educational Testing Service, Princeton, New Jersey*

**1. Introduction.** This note is concerned with the situation where $U$ fallible measurements of some single characteristic are made on each of a large number of objects. The $U$ measurements may represent $U$ different methods of measuring the same characteristic, each method involving a different frequency distribution of errors of measurement.

For each object, there is an unknown "true value" of the characteristic. The difference between the observed measurement and the true value is an error of measurement. The true value and the errors of measurement will be termed *latent variables*.

The results derived are currently being applied in psychometric work, but they should be applicable in almost any field where unbiased fallible measurements are made. For example, the true amount ($\xi$) of some chemical constituent of the blood may have been fallibly but independently measured by $U$ different methods (or by $U$ different laboratory technicians) for each of a large number of hospital patients. The results given here will permit the consistent estimation of the first $U$ cumulants of $\xi$, the first $U$ cumulants of the error of measurement in each of the $U$ methods, and, further, all the multivariate cumulants of the latent variables up through order $U$.

In psychometric work, $U$ strictly parallel forms of a mental test may be prepared by matching the questions assigned to the different forms on their statistical characteristics (determined by pretesting). These test forms may, in effect, all be administered "simultaneously" by the device of interspersing the questions from all forms and then scoring the questions of each form separately, counting the number answered correctly. The moments of the frequency distribution of the "true scores" ($\xi$) of the examinees tested and also the distribution of the errors of measurement may now be estimated by the method to be described. (In this case, the shape of the distribution of errors of measurement must be dependent on the value of $\xi$. This is apparent, for example, from the fact that the observed test scores cannot be negative; hence whenever $\xi$ is near zero large negative errors of measurement cannot occur.)

Formulas illustrating the final results obtained are given in Section 2. The derivations are given in the remaining sections.

In Section 3, any multivariate cumulant of the observed measurements is expressed as a linear function of the cumulants of the joint distribution of the latent variables, no assumption being made other than the existence of the cumulants in question (the results of this section are not new; they could be di-

Received August 29, 1958; revised December 1, 1958.

rectly obtained by specializing a formula given by James ([1], eq. 13), for example). In Section 4 it is assumed that each error of measurement has a mean of zero and is uncorrelated with other appropriate chance variables; it is shown that each multivariate (and univariate) cumulant of the latent variables can be expressed in closed form as a simple linear function of the cumulants of the observed measurements. Section 5 details the restrictions imposed on the cumulants of the observed measurements by the assumption about uncorrelated errors.

**2. Specific formulas.** There are $U$ *observed measurements* on each object, denoted by $x_1, \cdots, x_U$. Both these and the *true value*, $\xi$, are random variables. The *errors of measurement* are, by definition,

$$e_u = x_u - \xi, \qquad (u = 1, \cdots, U). \tag{1}$$

Any $U$-variate cumulant of the observed measurements is denoted by $\kappa_{C_1 \cdots C_U}$, where $C_1, \cdots, C_U$ are nonnegative integers referring to the variables $x_1, \cdots, x_U$, respectively. Any cumulant of the latent variables is similarly denoted by $K_{B_0, B_1 B_2 \cdots B_U}$, where the first subscript refers to variable $\xi$ and the other $U$ subscripts refer to the $U$ errors of measurement. It will be notationally convenient to use a zero-order cumulant, having all zero subscripts, that is by definition equal to zero.

Explicit formulas are given below expressing all latent-variable cumulants up through the fourth order in terms of the observed-variable cumulants for the case where $U = 4$. All necessary formulas are either given or may be obtained by permutation of subscripts. The first subscript on each $K$, representing the true value, is not subject to permutation, but the $U$ other subscripts on each $K$ or $\kappa$ may be permuted providing the same permutation is made on each $K$ and $\kappa$ throughout the entire formula.

$$K_{0,1000} = K_{0,0100} = K_{0,0010} = K_{0,0001} = 0 \text{ by assumption}$$

$$K_{1,0000} = \kappa_{1000} = \kappa_{0100} = \kappa_{0010} = \kappa_{0001}$$

$$K_{2,0000} = \kappa_{1100} = \kappa_{1010} = \cdots = \kappa_{0011}$$

$$K_{0,2000} = \kappa_{2000} - \kappa_{1100} = \cdots = \kappa_{2000} - \kappa_{0011}$$

$$K_{0,0200} = \kappa_{0200} - \kappa_{1100} = \cdots = \kappa_{0200} - \kappa_{0011}, \text{ etc.}$$

$$K_{3,0000} = \kappa_{1110} = \cdots = \kappa_{0111}$$

$$K_{2,1000} = \cdots = K_{2,0001} = 0, \text{ by assumption}$$

$$K_{1,1100} = K_{0,2100} = K_{0,1110} = 0, \text{ etc., by assumption}$$

$$K_{1,2000} = \kappa_{2100} - \kappa_{1110} = \cdots = \kappa_{2001} - \kappa_{0111}, \text{ etc.}$$

$$K_{0,3000} = \kappa_{3000} - 3\kappa_{2100} + 2\kappa_{1110}, \text{ etc.}$$

$$K_{4,0000} = \kappa_{1111}$$

$$K_{3,1000} = K_{2,1100} = K_{1,2100} = \cdots = K_{0,1111} = 0 \text{ by assumption}$$

$$K_{2,2000} = \kappa_{2110} - \kappa_{1111}, \text{ etc.}$$

$$K_{1,3000} = \kappa_{3100} - 3\kappa_{2110} + 2\kappa_{1111}, \text{ etc.}$$

$$K_{0,2200} = \kappa_{2200} - \kappa_{2110} - \kappa_{1210} + \kappa_{1111}, \text{ etc.}$$

$$K_{0,4000} = \kappa_{4000} - 4\kappa_{3100} + 6\kappa_{2110} - 3\kappa_{1111}, \text{ etc.}$$

**3. General relations among cumulants.** The characteristic function of the latent variables may be written

$$F(T_0, T_1, \cdots, T_U) = E \exp i \sum_{u=0}^{U} T_u e_u, \tag{2}$$

where $E$ is the expectation symbol and $e_0 = \xi$. That of the observed measurements is, by (1),

$$f(t_1, \cdots, t_U) = E \exp i \sum_{u=1}^{U} t_u x_u = E \exp i \sum_{u=0}^{U} t_u e_u, \tag{3}$$

where $t_0 = \sum_{u=1}^{U} t_u$.

It is seen that the first characteristic function may be changed to the second simply by replacing $T$ by $t$. If the necessary cumulants exist, the cumulant-generating function of the latent variables is

$$\log F(T_0, T_1, \cdots, T_U) = \sum_{B_0=0}^{\infty} \sum_{B_1=0}^{\infty} \cdots \sum_{B_U=0}^{\infty} \frac{i^P T_0^{B_0} T_1^{B_1} \cdots T_U^{B_U}}{B_0! B_1! \cdots B_U!} K_{B_0, B_1 \cdots B_U}, \tag{4}$$

where $P = \sum_{u=0}^{U} B_u$. Take the right side of (4) and replace $T_1, \cdots, T_U$ by $t_1, \cdots, t_U$ and $T_0^{B_0}$ by the multinomial expansion

$$t_0^{B_0} = \left(\sum_{u=1}^{U} t_u\right)^{B_0} \equiv \sum_a \frac{B_0!}{a_1! \cdots a_U!} t_1^{a_1} \cdots t_U^{a_U}, \tag{5}$$

where $a_1, \cdots, a_U$ are nonnegative integers and $\Sigma_a$ is over all sets of $a$ such that $\sum_{u=1}^{U} a_u = B_0$. This converts the right side of (4) to the cumulant-generating function of the observed measurements:

$$\log f(t_1, \cdots, t_U) = \sum_{B_0=0}^{\infty} \sum_{B_1=0}^{\infty} \cdots \sum_{B_U=0}^{\infty} \sum_a \frac{i^P t_1^{B_1+a_1} \cdots t_U^{B_U+a_U}}{B_1! \cdots B_U! a_1! \cdots a_U!} K_{B_0, B_1 \cdots B_U}. \tag{6}$$

The cumulant $\kappa_{C_1 \cdots C_U}$ of the observed measurements is the coefficient of the term $i^{P'} t_1^{C_1} \cdots t_U^{C_U}/C_1! \cdots C_U!$ in the series at the right, where $P' = \sum_{u=1}^{U} C_u$. If $B_u + a_u$ is replaced by $C_u$ in (6) and the terms rearranged, these cumulants are found to be

$$\kappa_{C_1 \cdots C_U} = \sum_B K_{B_0, B_1 \cdots B_U} \prod_{u=1}^{U} \binom{C_u}{B_u}, \tag{7}$$

where $\binom{C_u}{B_u} = C_u!/B_u!(C_u - B_u)!$, and where $\Sigma_B$ is taken over all sets of nonnegative integral values of the $B_u$ subject to the restrictions that $B_u \leqq C_u$ for $u = 1, \cdots, U$ and $P' = P$.

The result given in equation (7) may also be expressed in terms of symbolic multiplication:

$$\kappa_{C_1 \cdots C_U} \sim (\xi + e_1)^{C_1}(\xi + e_2)^{C_2} \cdots (\xi + e_U)^{C_U}. \tag{8}$$

The $\sim$ symbol may be replaced by an equals sign when each term $\xi^{B_0}e_1^{B_1}e_2^{B_2} \cdots e_U^{B_U}$ on the right has been replaced by $K_{B_0, B_1 \cdots B_U}$.[1]

Formulas (7) or (8) express any $P$th-order cumulant of the observed measurements as a linear function of the $P$th-order cumulants of the latent variables. Without further assumptions, it is not possible to solve any set of these equations for the unknown cumulants of the latent variables, for the reason that there are always fewer equations than unknowns.

**4. Determining the cumulants of the latent variables.** The result in (7) and (8) was obtained without any assumption about the distribution of the latent variables other than the existence of the cumulants. It will now be assumed that each error of measurement has a mean value of 0 and is uncorrelated with every product of the remaining latent variables. Thus $K_{B_0, B_1 \cdots B_U} = 0$ whenever any $B_u$ $(u > 0)$ is equal to 1. This is much less restrictive than the usual assumption that the errors of measurement are distributed independently of the true value and of each other. *The present assumption, for example, permits the variance of the errors of measurement and all higher moments to be dependent on* $\xi$—it is only the mean error of measurement that is independent of $\xi$.

With this assumption, we may proceed to prove

THEOREM 1. *Given that* $K_{B_0, B_1 \cdots B_U} = 0$ *whenever any* $B_u = 1$ $(u > 0)$, *all equations* (7) *for which* $\sum_{u=1}^{U} C_u = P$ *is constant can be ranked so that the right side of each contains at most one nonzero K appearing in no preceding equation; thus, given that the equations are consistent, they may be solved so as to express any K of order* $\leqq U$ *as a linear function of* $\kappa$*'s.*

Let $U - T$ be the number of subscripts on the left side of (7) that are equal to 1, so that the observed-variable cumulant may be written $\kappa_{C_1 C_2 \cdots C_T 11 \cdots 1}$. For $B_0 < U - T$, there must be at least one value of $u > 0$ for which $B_u = 1$ on the right side of (7), so every $K_{B_0, B_1 \cdots B_U}$ will vanish whenever $B_0 < U - T$. For $B_0 = U - T$, there is on the right side of (7) one and only one cumulant, $K_{B_0, B_1 \cdots B_U}$, without unit subscripts; this unique nonvanishing cumulant has $T$ subscripts that are the same as those of the observed-variable cumulant and (at least) $U - T$ zero subscripts, so it may be written $K_{(U-T), C_1 C_2 \cdots C_T 00 \cdots 0}$; it will be spoken of as the latent-variable cumulant to which $\kappa_{C_1 C_2 \cdots C_T 11 \cdots 1}$ (with

[1] As pointed out by a referee, formula (8) shows that the relation between observed-variable and latent-variable cumulants is exactly the same as the relation between observed-variable and latent-variable moments about an arbitrary origin.

$U - T$ unit subscripts) *corresponds.* For $B_0 > U - T$, there may be on the right side of (7) a number of cumulants $K_{B_0,B_1\cdots B_U}$ without unit subscripts; each of these, however, is a latent-variable cumulant to which some other observed-variable cumulant $\kappa_{B_1B_2\cdots B_U11\cdots1}$ having $B_0$ unit subscripts corresponds.

Consider all the observed-variable cumulants of a given order ($P$) and suppose them to be grouped according to $U - T$, the number of unit subscripts, and the groups ranked on $U - T$ starting with the cumulant for which $T = 0$. The results of the preceding paragraph show that equation (7) expresses any $\kappa$ as a linear function of $K$'s, one of these being the $K$ to which the given $\kappa$ corresponds, all the others being $K$'s to which correspond some other $\kappa$'s of lower rank. This completes the proof of Theorem 1.

**5. Restrictions on the observed cumulants.** If none of the $C$'s are zero in $\kappa_{C_1C_2\cdots C_T11\cdots1}$, then this is the only $\kappa$ that corresponds to $K_{(U-T),C_1C_2\cdots C_T00\cdots0}$. If $V$ of the $C$'s in $\kappa_{C_1C_2\cdots C_T11\cdots1}$ are zero, then every $\kappa$ obtained by permuting the zero and unit subscripts on $\kappa$ also corresponds to the same $K$.

It follows from Theorem 1 that any $K$ may be expressed as equal to any one of the $\binom{V+U-T}{U-T}$ corresponding $\kappa$'s plus other $\kappa$'s of lower rank (unless the first $\kappa$ already is of lowest rank). For the $\kappa$'s of lowest rank, $T = V$ and there are $\binom{U}{U-V}$ equally good equations, such as $K_{(U-T),00\cdots0} = \kappa_{11\cdots100\cdots0}$, there being $U$ zero subscripts on the left side of the equation, $T$ zero subscripts and $U - T$ unit subscripts on the right. These $\binom{U}{U-T}$ different $\kappa$'s must thus be equal. Proceeding to the case of next higher rank, another set of $\kappa$'s are found that must be equal to each other. Mathematical induction now shows that all $\kappa$'s corresponding to a given $K$ must be equal. Thus,

THEOREM 2. *Given that $K_{B_0,B_1\cdots B_U} = 0$ for any $B_u = 1$ $(u > 0)$, any two $\kappa$'s will be equal if their subscripts are the same except for a permutation that involves zero and unit subscripts only.*

Theorem 2 states a restriction on the observed-variable $\kappa$'s that is implicit in the assumption made in Section 4 about uncorrelated errors. Since the matrix of the large-sample sampling variances and covariances of the $\kappa$'s could be computed if desired, the assumption made in Section 4 can be submitted to statistical test, at least in large samples, to determine whether or not any given set of observed data is compatible with it.

REFERENCE

[1] G. S. JAMES, "Cumulants of a transformed variate," *Biometrika*, Vol. 42 (1955), pp. 529–531.

# SOME TESTS OF PERMUTATION SYMMETRY[1]

By R. Wormleighton

*University of Toronto*

**Summary.** The two-sample sign test is viewed as a test for the permutation symmetry of a bivariate distribution, and extensions to $k$-variate distributions are sought. Friedman's rank test, [1], although originally intended as a substitute for the $F$-test in a two-way classification, is such an extension. Study of the family of two-sample sign tests obtained by comparing the $k$ coordinates pairwise has yielded a statistic with an asymptotic Chi-square distribution from which a further test of symmetry can be constructed. The statistic is based on more degrees of freedom than Friedman's and is sensitive to a greater variety of alternatives. This extension is analogous to that obtained by Terpstra [2] from the Wilcoxon test.[2] In this case, however, the limiting distribution turns out to be non-singular. The argument leading to the test is not restricted to the case of complete symmetry but may be carried through with any specified degree of asymmetry. The coordinates may also be compared $m$ at a time, $2 \leqq m \leqq k$. The argument can be extended and, with a slight modification, includes the derivation of Friedman's test. Thus a hierarchy of tests of permutation symmetry are available: Friedman's test corresponds to the case, $m = 1$; when $m = k$, the corresponding test turns out to be Pearson's Chi-square.

**1. Introduction.** Given $n$ pairs of observations, ... sometimes called two "matched" samples, ... the sign test statistic for comparing the populations from which the two matched samples were drawn is the number of cases in which the first observation of a pair is greater than the second; a simple count. The statistic and its distribution are easily computed, the test requires minimal assumptions about the underlying probability distributions, and when these distributions are normal, the efficiency of the test relative to the $t$-test is high [3]. It is natural, therefore, to explore extensions of the test to three or more matched samples.

In the case of three samples, J. W. Tukey has suggested the following approximate test. To make the test at level $\alpha'$, conduct ordinary two-tailed sign tests comparing each of the three pairs of samples, at level $\alpha = \alpha'/3$. If one or more of these three tests yields a significant result, the combined test is significant. This is a convenient approximation, but it does not appear to be worth while to extend this method to the case of more than three samples.

In the general case of $k$ matched samples of $n$—that is to say, $n$ observations on $k$-variate distributions—one extension has been given by Friedman [1]. The

Received April 16, 1956; revised April 13, 1959.

[1] Work done in part under a grant from the Office of Naval Research held at Princeton University.

[2] The author is indebted to the editors for drawing his attention to the work of Terpstra.

$k$ coordinates of each observation are ranked in order of magnitude, and the mean ranks calculated. The sum of squares of deviations of these $k$ mean ranks from their general mean is proportional to a statistic $X_r^2$ which, under the appropriate null hypothesis, has asymptotically a Chi-square distribution with $k - 1$ degrees of freedom. Another extension is suggested in the preceding paragraph. Consider the family of two-sample sign tests obtained by comparing the $k$ coordinates pairwise; the corresponding statistics form a set of $C_2^k$ simple counts, the number of times the $i$th coordinate exceeds the $j$th in the sample. This paper is primarily concerned with this set of simple counts.

Underlying much discussion of the sign test is an intuitive picture of two independent and similar, if not identical, populations which are to be compared for differences in location. Friedman's rank test is the natural generalization. However, regarding the two paired samples as a single sample from a bivariate distribution, the sign test becomes a test of the permutation symmetry of the distribution; the null hypothesis states that both orderings of the coordinates are equally probable. In the case of a $k$-variate distribution, the most general non-parametric test of permutation symmetry is based on the statistic (of dimension, $k!$) giving the number of times each of the $k$ possible orderings occurs in the sample. While such a test is of use against any non-symmetric alternative, it would appear that, unless the sample size is of the order $k!$, only the most extreme departures from symmetry would be detected. If only certain kinds of asymmetry are of interest then a more specific test is required; Friedman's rank is an example of such a specific test.

Suppose one wishes to determine whether a card-shuffling device is acceptable. Ideally, no matter what the order of the cards before shuffling, all orderings should be equally probable after shuffling. It would be acceptable, however, if it were practically impossible for a card-player, knowing the initial order of the cards, and the final position of some of them, to draw inferences about the position of the remainder. A bridge player holding the Queen of Spades should not be able to infer from the previous hand that the King of Spades is more likely to be on his right than on his left. If the initial position of a card is $i$, let $X_i$ denote its final position after shuffling. Then we are concerned with the symmetry of the multivariate distribution of the $\{X_i\}$. The most general test may very well require an impossibly large experiment—52! is a very large number. On the other hand, Friedman's rank test is too specific. A single cut, provided all possible places for the cut are equally likely, (including no cut at all), is sufficient to ensure that the expectation of $X_i$ is the same for all $i$; and, as is shown below, this implies that the expectation of $X_r^2$ is the same as it would be under the null hypothesis of complete symmetry. Clearly, something intermediate is required.

**2. Notation.** Let

$$X^{(a)} = (X_1^{(a)}, X_2^{(a)}, \cdots, X_k^{(a)}), \qquad a = 1, 2, \cdots, n,$$

represent $n$ $k$-variate real-valued random variables. Assume, as null hypothesis, that

$$\Pr\{X_{i_1}^{(a)} < X_{i_2}^{(a)} < \cdots < X_{i_k}^{(a)}\} = 1/k! \tag{2.1}$$

for all permutations $(i_1, i_2, \cdots, i_k)$ of the subscripts $(1, 2, \cdots, k)$, and all $a$.

Let

$$\begin{aligned} Y_{ij}^{(a)} &= 1 \text{ if } X_i^{(a)} > X_j^{(a)} \qquad i \neq j \\ &= 0 \text{ otherwise,} \end{aligned} \tag{2.2}$$

and define

$$\beta_{ij} = \frac{2}{\sqrt{n}} \sum_{a=1}^{n} (Y_{ij}^{(a)} - \tfrac{1}{2}). \tag{2.3}$$

Since $\beta_{ij} + \beta_{ji} = 0$, we need consider only one of each pair, $\{\beta_{ij}, \beta_{ji}\}$. It will appear (Lemma 1), that the choice does not affect our conclusions and we consider, therefore, only the set $\{\beta_{ij}\}$ for which $i < j$.

The exact probability at any point $\{\beta_{ij}\}$ is given by a sum of multinomial terms.

THEOREM I:

$$\Pr(\beta_{ij}) = \frac{1}{(k!)^n} \sum \frac{n!}{\prod_{h=1}^{k!} \alpha_h} \tag{2.4}$$

*where the sum is taken over all sets of non-negative integers* $\{\alpha_h\}$, $(h = 1, 2, \cdots, k!)$, *satisfying*

$$\sum_{h=1}^{k!} \alpha_h = n \tag{2.5}$$

*and a set of* $C_2^k$ *linear equations of the form*

$$2 \sum_{h=1}^{k!} c_{ijh}\alpha_h = \sqrt{n}\beta_{ij} - n, \qquad c_{ijh} = 1 \text{ or } 0 \text{ as required.} \tag{2.6}$$

The moments of the $\beta_{ij}$ can be computed directly. $E\{\beta_{ij}\} = 0$. For any admissible choice of $\{\beta_{ij}\}$, the covariance matrix is non-singular and can be inverted.

THEOREM II: *Let* $(\sigma_{ii',jj'})$ *be the covariance matrix of the* $C_2^k$*-dimensional random variate* $\{\beta_{ij}\}$. *Let D be its determinant and let* $(\sigma^{ii',jj'})$ *be its inverse. Then for all permissible choices of the coordinates* $\{\beta_{ij}\}$,

$$D = (k+1)^{k-1}/3^{\frac{k(k-1)}{2}}; \tag{2.7}$$

| | | | | |
|---|---|---|---|---|
| *if* | $i = j$, | $i' = j'$, | $\sigma_{ii',jj'} = 1$ | $\sigma^{ii',jj'} = \dfrac{3(k-1)}{k+1}$, |
| *if* | $i = j'$, | $i' = j$, | $= -1$ | not defined |
| *if* | $i = j$, | $i' \neq j'$, | | |
| *or* | $i \neq j$, | $i' = j'$, | $= \frac{1}{3}$ | $= \dfrac{-3}{k+1}$ |

$$
\begin{array}{llll}
\textit{if} & i = j', \quad i' \neq j, & & \\
\textit{or} & i \neq j', \quad i' = j, & = -\tfrac{1}{3} & = \dfrac{3}{k+1} \\
\textit{otherwise,} & & = 0 & = 0.
\end{array}
$$

Outline of proof: In the determinant, $D$, rows can be replaced by linear combinations of rows to introduce blocks of zeros. Precisely, we replace $\sigma_{ii',jp}$ by

$$\tau_{ii',jp} = \sigma_{ii',jp} - \frac{1}{p} \sum_{\substack{\nu \\ (\nu<p)}} \sigma_{ii',j\nu}, \qquad \text{all } (ii'). \tag{2.8}$$

$D$ can then be written as a product of principal minors,

$$D = \prod_{p=2}^{k} d_p, \tag{2.9}$$

where $d_p$ is of order $(p - 1)$ with elements $(2(p + 1))/3p$ on the diagonal, and $(p + 1)/3p$, elsewhere.

$$d_p = p\left[\frac{p+1}{3p}\right]^{p-1}. \tag{2.10}$$

Hence, $D = (k + 1)^{k-1}/3^{(k(k-1))/2}$.

To verify the remainder of the theorem, we multiply the covariance matrix by its stated inverse, and evaluate the sums, $\sum_{(ii')}\sigma_{hh',ii'}\sigma^{ii',jj'}$. A term of such a sum will differ from zero only if the pair of indices $(ii')$ has an element in common with each of the pairs, $(hh')$ and $(jj')$.

For a diagonal element of the product, $(hh') = (jj')$. There is one term in the sum with $(ii') = (hh')$ of value, $(3(k - 1))/(k + 1)$; and $2(k - 2)$ terms in which $(ii')$ has one element in common with $(hh')$, each of value $-1/(k + 1)$; all other terms are zero, and the sum is unity.

For non-diagonal elements of the product, where $(hh')$ and $(jj')$ have one or zero elements in common, we have, for example:

(i) $\sum_{(ii')} \sigma_{12,ii'}\sigma^{ii',13} = \sigma_{12,12}\sigma^{12,13} + \sigma_{12,13}\sigma^{13,13} + \sigma_{12,23}\sigma^{23,13} + \sum_{i=4}^{k} \sigma_{12,1i'}\sigma^{1i',13} = 0$

(ii) $\sum_{(ii')} \sigma_{12,ii'}\sigma^{ii',34} = \sigma_{12,13}\sigma^{13,34} + \sigma_{12,14}\sigma^{14,34} + \sigma_{12,23}\sigma^{23,34} + \sigma_{12,24}\sigma^{24,34} = 0.$

**3. The statistic $\chi_t^2$.** We now define the statistic on which our first test of permutation symmetry is based. Let

$$\chi_t^2 = \sum_{i<i'} \sum_{j<j'} \sigma^{ii',jj'} \beta_{ii'}\beta_{jj'} \tag{3.1}$$

$\chi_t^2$ is thus the quadratic form associated with the inverse of the covariance matrix of $\beta$, for a particular choice of the coordinates of $\beta$. This particular choice is only

TABLE I

*Distribution of $\chi_i^2$ for $k = 3$; $n = 2, 3, 4, 5, 6$*

| $x$ | $Pr\ (\chi_i^2 \leq x)$ | $x$ | $Pr\ (\chi_i^2 \leq x)$ | $x$ | $Pr\ (\chi_i^2 \leq x)$ |
|---|---|---|---|---|---|
| $n = 2$ | | $n = 4$ | | $n = 6$ | |
| 0 | .167 | 0 | .0694 | 0 | .0399 |
| 1.5 | .833 | 1.5 | .5139 | 1.0 | .3485 |
| 3.0 | 1.000 | 3.0 | .6435 | 2.0 | .4507 |
| | | 4.5 | .8657 | 3.0 | .6977 |
| $n = 3$ | | 6.0 | .9193 | 4.0 | .7748 |
| 1.0 | .417 | 7.5 | .9954 | 5.0 | .8828 |
| 3.0 | .639 | 12.0 | 1.0000 | 6.0 | .8983 |
| 5.0 | .972 | | | 7.0 | .9600 |
| 9.0 | 1.000 | $n = 5$ | | 8.0 | .9685 |
| | | 0.6 | .2392 | 9.0 | .9891 |
| | | 1.8 | .4244 | 10.0 | .9968 |
| | | 3.0 | .7639 | 13.0 | .9999 |
| | | 5.4 | .8912 | 18.0 | 1.0000 |
| | | 6.6 | .9529 | | |
| | | 7.8 | .9838 | | |
| | | 10.2 | .9992 | | |
| | | 15.0 | 1.0000 | | |

a notational convenience; the same statistic is obtained with any other admissible choice, (Lemma 1).

The exact distribution of $\chi_i^2$, for $n$ finite, can be computed from the distribution of $\beta$ as given in Theorem I. This is an arduous process, except when $k$ and $n$ are both small. A few values are given in Table I, for $k = 3$.

The asymptotic distributions of $\beta$ and $\chi_i^2$ are given by

THEOREM III: *The vector random variate, $\beta = \{\beta_{ij}\}$, has asymptotically a non-singular multivariate normal distribution with density function $Ce^{-\frac{1}{2}\chi_i^2}$. The statistic $\chi_i^2$ is asymptotically distributed as Chi-square with $C_2^k$ degrees of freedom.*

PROOF: $\beta$ is the standardized sum of $n$ identically and independently distributed vector random variates. Since all second moments are finite, the simplest conditions for the central limit theorem in its multivariate form, [4], are satisfied. Therefore, as $n$ increases, the distribution of $\beta$ tends to the multivariate normal. The covariance matrix is non-singular and independent of $n$; hence, the density function exists.

It is well-known that the exponent of the density function is distributed as Chi-square, [5]. Hence, $\chi_i^2$ is asymptotically Chi-square with $C_2^k$ degrees of freedom.

The null hypothesis should be rejected whenever $\chi_i^2$ is large.

An indication of the accuracy of the asymptotic approximation is given in Table II for $k = 3$ and small $n$.

TABLE II

*Comparison of $X_t^2$ with approximating Chi-square, $X_3^2$, with 3 d.f. ($k = 3$)*

| $x$ | $Pr(X_t^2 \geq x)$ | $Pr(X_3^2 \geq x)$ | $x$ | $Pr(X_t^2 \geq x)$ | $Pr(X_3^2 \geq x)$ |
|---|---|---|---|---|---|
| | $n = 3$ | | | $n = 5$ | |
| 9.0 | .028 | .029 | 10.2 | .0162 | .0170 |
| 5.0 | .361 | .172* | 7.8 | .0471 | .0504 |
| | | | 6.6 | .1088 | .0859* |
| | $n = 4$ | | | $n = 6$ | |
| 12.0 | .0046 | .0074 | 9.0 | .0315 | .0295 |
| 7.5 | .0807 | .0576* | 8.0 | .0400 | .0461 |
| 6.0 | .1343 | .1117 | 7.0 | .1017 | .0720* |

* Only in the cases marked by an asterisk is the approximation improved by a continuity correction.

**4. Friedman's rank test.** There exists a hierarchy of tests of permutation symmetry, one of which is the $X_t^2$-test of the previous section. Another such test, lying at one end of the chain, is Friedman's rank test [1].

In the rank test, the $k$ coordinates of an observation are ranked in increasing order of magnitude. If $r_{i\nu}$ denotes the rank of coordinate $X_i$, in the $\nu$th observation, then $r_{i\nu} - 1 =$ no. of coordinates less than $X_i$. Let

$$\bar{\rho}_i = \frac{1}{n} \sum_{\nu=1}^{n} \left( r_{i\nu} - \frac{k+1}{2} \right). \tag{4.1}$$

Friedman proposed the statistic

$$\chi_r^2 = \frac{12n}{k(k+1)} \sum_{i=1}^{k} \bar{\rho}_i^2 \tag{4.2}$$

for testing the null hypothesis. Large values of $\chi_r^2$ lead to rejection of the hypothesis.

Friedman tabulated the exact distribution of $\chi_r^2$ for small $n$ and $k$, and gave a proof that $\chi_r^2$ is asymptotically distributed as Chi-square with $(k - 1)$ degrees of freedom.

We note that

$$\bar{\rho}_i = \frac{1}{2\sqrt{n}} \sum_{\substack{j \\ (j \neq i)}} \beta_{ij}\,. \tag{4.3}$$

Hence,

$$\chi_r^2 = \frac{3}{k(k+1)} \sum_{i=1}^{k} \left[ \sum_{\substack{j \\ (j \neq i)}} \beta_{ij} \right]^2 \tag{4.4}$$

and

$$\chi_t^2 = 3 \sum_{\substack{ij \\ (i<j)}} \beta_{ij}^2 - k\chi_r^2 \tag{4.5}$$

We also define

$$\chi_\Delta^2 = \chi_t^2 - \chi_r^2 . \tag{4.6}$$

From Fisher's Lemma [6], we can deduce that $\chi_\Delta^2$ is asymptotically Chi-square with $C_2^k - (k - 1) = C_2^{k-1}$ degrees of freedom, and independent of $\chi_r^2$.

Friedman also showed that, for finite $n$,

$$E\chi_r^2 = k - 1 \qquad \text{Var } \chi_r^2 = \frac{n-1}{n} 2(k-1). \tag{4.8}$$

Similarly, we obtain, by direct calculation,

$$\begin{aligned} E\chi_t^2 &= C_2^k & \text{Var } \chi_t^2 &= \frac{n-1}{n} k(k-1) \\ E\chi_\Delta^2 &= C_2^{k-1} & \text{Var } \chi_\Delta^2 &= \frac{n-1}{n} (k-1)(k-2). \end{aligned} \tag{4.9}$$

For each of these statistics, the mean is independent of $n$, and the variance is an increasing function of $n$. This strongly suggests that the use of the asymptotic distribution, (without a continuity correction), for defining the critical region will lead to errors in the so-called "safe" direction; i.e., the true size of the critical region will be smaller than the significance level. The computations carried out by Friedman, and by the writer, support this.

$\chi_r^2$ and $\chi_t^2$ provides tests of the same null hypothesis. In most practical applications, the alternatives of interest—e.g., one or more coordinates tending to be consistently higher than the remainder—can be distinguished by either test. In the writer's opinion, $\chi_r^2$ is usually the preferred test, and $\chi_t^2$ should be reserved for special situations. An experimenter may, on occasion, wish to make two tests based on $\chi_r^2$ and $\chi_\Delta^2$. These two tests are asymptotically independent and are sensitive to two distinct classes of alternatives. The classification of alternatives is discussed in Section 9.

**5. First extension: arbitrary null hypothesis.** Although the argument has been presented with one particular hypothesis as null hypothesis—viz., all orderings of the $\{X_i\}$ equally likely—it could just as easily be carried through with an asymmetric null hypothesis:

$$\Pr\{X_{i_1} > X_{i_2} > \cdots > X_{i_k}\} = p_{i_1 i_2 \cdots i_k} > 0 \tag{5.1}$$

where $p_{i_1 i_2 \cdots i_k}$ is a set of $k!$ positive numbers which sum to unity.

As a notational convenience, we introduce symbols representing sums of the constants in the null hypothesis.

$$p_{i_1 i_2 \cdots i_m} = \Pr\{X_{i_1} > X_{i_2} > \cdots > X_{i_m}\}, \qquad 2 \leq m < k. \tag{5.2}$$

Random variables $Y_{ij}^{(\alpha)}$ are defined as before, (2.2), and the standardized variates, $\beta_{ij}$, are defined similarly by

$$\beta_{ij} = \frac{1}{\sqrt{np_{ij}(1 - p_{ij})}} \sum_{\alpha=1}^{n} (Y_{ij}^{(\alpha)} - p_{ij}). \tag{5.3}$$

The vector variate, $\beta$, is defined by $C_2^k$ coordinates, $\beta_{ij}$, which satisfy no linear relation.

Let $\sigma_{ii',jj'} = \mathrm{Covar}\{\beta_{ii'}, \beta_{jj'}\}$. As before, $\mathrm{Var}(\beta_{ii'}) = 1$, and when $i, i', j, j'$ are all distinct, $\sigma_{ii',jj'} = 0$. However, the covariance of two coordinates with one subscript in common is more complex; e.g.

$$\begin{aligned} \sigma_{12,13} = {} & \sqrt{\frac{(1 - p_{12})(1 - p_{13})}{p_{12}\, p_{13}}} \cdot (p_{123} + p_{132}) - \sqrt{\frac{(1 - p_{12})p_{13}}{p_{12}(1 - p_{13})}} \cdot p_{312} \\ & - \sqrt{\frac{p_{12}(1 - p_{13})}{(1 - p_{12})p_{13}}} \cdot p_{213} + \sqrt{\frac{p_{12}\, p_{13}}{(1 - p_{12})(1 - p_{13})}} \cdot (p_{231} + p_{321}). \end{aligned} \tag{5.4}$$

However, all variances and covariances are finite, and the central limit theorem still applies; therefore, $\beta$ has asymptotically a multivariate normal distribution.

It will be shown (Theorem IV) that the rank of the matrix $(\sigma_{ii',jj'})$ of order $C_2^k$, is also $C_2^k$; hence, its inverse $(\sigma^{ii',jj'})$ exists. We can therefore define

$$Q_2 = \sum_{\substack{ii' \\ (i<i')}} \sum_{\substack{jj \\ (j<j')}} \sigma^{ii',jj'} \beta_{ii'} \beta_{jj'}. \tag{5.5}$$

$Q_2$ is the exponent of the density function of the asymptotic distribution and we deduce, as before, that $Q_2$ is asymptotically Chi-square with $C_2^k$ degrees of freedom.

To test the null hypothesis, reject when $Q_2$ is large. The critical region can be determined easily, and approximately, with the asymptotic distribution.

**6. Second extension: the $Q_m$-statistics.** Instead of comparing the random variables, $\{X_i\}$, two at a time, we could compare them $m$ at a time ($2 \leq m \leq k$).

We define new random variables of order $m$, by

$$\begin{aligned} Z_{i_1 i_2 \cdots i_m}^{(\alpha)} &= 1 \text{ if } X_{i_1}^{(\alpha)} > X_{i_2}^{(\alpha)} > \cdots > X_{i_m}^{(\alpha)} \\ &= 0 \text{ otherwise} \end{aligned} \tag{6.1}$$

where $(i_1, i_2, \cdots, i_m)$ is a subset of the first $k$ positive integers.

There are $k(k - 1) \cdots (k - m + 1) = \theta_m$, say, such random variables of order $m$.

If $n$ observations are made, then $\sum_{\alpha=1}^{n} Z_{i_1 i_2 \cdots i_m}^{(\alpha)} =$ the number of times in which the ordering $X_{i_1} > X_{i_2} > \cdots > X_{i_m}$ occurs.

We define the standardized random variate of order $m$ by

$$\gamma_{i_1 i_2 \cdots i_m} = \frac{1}{\sqrt{np_{i_1 i_2 \cdots i_m}(1 - p_{i_1 \cdots i_m})}} \sum_{\alpha=1}^{n} (Z_{i_1 i_2 \cdots i_m}^{(\alpha)} - p_{i_1 \cdots i_m}). \tag{6.2}$$

For any fixed set of integers $(i_1 i_2 \cdots i_m)$, the sum, over all permutations of the set, of the $Z^{(a)}_{i_1 i_2 \cdots i_m}$, is unity.

The standardized random variates of order $m$ satisfy $\tau_m = C^k_m$ independent linear relations, $(m \geq 2)$.

When $m = 2$, $\gamma_{i_1 i_2} = \beta_{i_1 i_2}$ as defined in Section 5.

When $m = 1$, we extend the definition by defining

$\gamma_i$ = standardized mean rank

$\tau_1 = 1$ (not $C^k_1$, since the $\gamma_i$ satisfy only one linear relation).

For every fixed $m$, a vector random variate $\gamma'_m = \{\gamma_{i_1 i_2 \cdots i_m}\}$ of dimension $\theta_m$, is defined.

By the central limit theorem, the distribution of $\gamma'_m$, approaches the multivariate normal distribution. However, because the coordinates satisfy $\tau_m$ linear relations the distribution is singular in the full $\theta_m$-space.

We reduce the space to dimension $\theta_m - \tau_m$ by omitting one of the coordinates appearing in each of the $\tau_m$ linear relations. (No coordinate appears in more than one such relation, so that exactly $\tau_m$ coordinates are omitted.) We denote the resulting vector random variate of dimension $(\theta_m - \tau_m)$ by $\gamma_m$. It will be shown in Theorem IV that the covariance matrix of $\gamma_m$ is non-singular and therefore, (at least in theory), can be inverted. Hence, we can define the statistic $Q_m$ = the quadratic form in $\gamma_{i_1 i_2 \cdots i_m}$ associated with the inverse of the reduced covariance matrix. $Q_m$ is asymptotically Chi-square with $(\theta_m - \tau_m)$ degrees of freedom.

LEMMA 1: *For fixed m, and a given simple null hypothesis, $Q_m$ is a uniquely defined function of the original sample.*

PROOF: Non-uniqueness could only occur when reducing the $\theta_m$-space to dimension $(\theta_m - \tau_m)$ by different choices of the coordinates to be retained.

Let $Q_m$ be defined in terms of one set $\gamma_{i_1 \cdots i_m}$ of coordinates, and let $\bar{Q}_m$ be defined in terms of a second set, $\bar{\gamma}_{i_1 \cdots i_m}$, of these coordinates. Using the known linear relations, each coordinate of the second set can be expressed as a linear combination of coordinates in the first set; thus, $\bar{Q}_m$ is also a quadratic form in the first set of coordinates. We must show, then, that corresponding coefficients in the two forms are equal. It is clearly sufficient to consider only the asymptotic distributions for, since the coefficients are independent of $n$, equality of coefficients in the limit implies equality for all $n$.

Consider, therefore, two quadratic forms, $Q$, $\bar{Q}$, in the random variates $U_1$, $U_2, \cdots, U_s$ where $U' = \{U_1, U_2, \cdots, U_s\}$ has an $s$-variate normal distribution and $Q = U'AU$, $\bar{Q} = U'\bar{A}U$, have Chi-square distributions with $s$ degrees of freedom. This implies that the forms $Q$, $\bar{Q}$ are of full rank and their associated matrices $A$, $\bar{A}$ are non-singular.

Therefore, there exist linear transformations

$$V = CU, \qquad \bar{V} = \bar{C}U \tag{6.3}$$

transforming $Q$, $\bar{Q}$ into sums of squares of $s$ independent standard normal variates, where $V$, $\bar{V}$ denote $s$-dimensional column vectors and $C$, $\bar{C}$ non-singular $s \times s$ matrices

$$Q = V'V, \qquad \tilde{Q} = \tilde{V}'\tilde{V}. \tag{6.4}$$

Clearly, $V = C\tilde{C}^{-1}\tilde{V} = P\tilde{V}$, say. Since the coordinates of both $V$ and $\tilde{V}$ are independent standard normal variates, $P$ is an orthogonal matrix. Thus $V'V = \tilde{V}'\tilde{V}$ and $Q = V'V = \tilde{V}'\tilde{V} = \tilde{Q}$.

Essential for the validity of the proof, and for the truth of the lemma, is the fact that both quadratic forms are of full rank. Otherwise, $C$ and $\tilde{C}$ are singular matrices without inverses, and $V$ is not, in general, a linear transform of $\tilde{V}$.

**7. Rank of $Q_m$: Expectation of $Q_m$.** We have already used the fact that the rank of the quadratic form $Q_m$ is $\theta_m - \tau_m$. We now give a proof of this statement.

THEOREM IV: *Let $\gamma_m$ be the vector random variate of dimension $(\theta_m - \tau_m)$ defined in Section 6 and let $A_m$ denote its covariance matrix of rank, $r_m$. Then $r_m = \theta_m - \tau_m$.*

LEMMA 2: *$r_m < \theta_m - \tau_m$ if, and only if, a linear relation holds among the coordinates of $\gamma_m$ with probability one.* Proof omitted.

LEMMA 3: *Each random variate $\gamma_{i_1 i_2 \cdots i_m}$ of order $m$ can be expressed as a linear combination of random variates of order $(m + 1)$.*

PROOF: $Z_{i_1 i_2 \cdots i_m} = 1$ whenever the ordering $X_{i_1} > X_{i_2} > \cdots > X_{i_m}$ occurs, i.e. whenever one of the orderings $(X_j > X_{i_1} > \cdots > X_{i_m})$, $(X_{i_1} > X_j > X_{i_2} > \cdots > X_{i_m})$, $\cdots$, $(X_{i_1} > \cdots > X_{i_m} > X_j)$ occurs, for fixed $j$ not a member of the set $(i_1, i_2, \cdots, i_m)$. $Z_{i_1 i_2 \cdots i_m} = Z_{j i_1 \cdots i_m} + Z_{i_1 j i_2 \cdots i_m} + \cdots + Z_{i_1 i_2 \cdots i_m j}$. The $\gamma$-variates are linear combinations of the $z$-variates of the same order. The lemma follows.

LEMMA 4: $r_k = k! - 1$.

PROOF: When $m = k$ there is only one choice for the set of integers $i_1 i_2 \cdots i_m$. Let $h$ index the permutations of this set. By direct calculations we obtain

$$\operatorname{Var}(\gamma_h) = 1, \qquad \operatorname{Covar}(\gamma_h, \gamma_{h'}) = -\sqrt{\frac{p_h p_{h'}}{(1 - p_h)(1 - p_{h'})}}. \tag{7.1}$$

We omit the coordinate corresponding to $h = k!$ to obtain $A_k$.

The determinant of $A_k$ can be evaluated directly.

$$|A_k| = \prod_{h=1}^{k!-1} \frac{p_1}{(1 - p_h)} \neq 0 \tag{7.2}$$

since by assumption, $p_1 \neq 0$

$$\therefore r_k = k! - 1. \tag{7.3}$$

PROOF OF THEOREM IV: Suppose $r_m < \theta_m - \tau_m$ for some $m$.

By Lemma 2 there exists a linear relation among the random variates $\gamma_{i_1 i_2 \cdots i_m}$ of order $m$ represented in $A_m$. But, by Lemma 3 each of these can be expressed as a linear combination of random variates of order $(m + 1)$ represented in $A_{m+1}$. Thus there exists a linear relation among the variates of order $(m + 1)$ and $r_{m+1} < \theta_{m+1} - \tau_{m+1}$. By induction, $r_k < k! - 1$. But this contradicts Lemma 4, hence $r_m = \theta_m - \tau_m$.

The expected value of the statistic, $Q_m$, for finite $n$, is given by

THEOREM V: $E\{Q_m\} = \theta_m - \tau_m$.

PROOF: We can write

$$Q_m = \sum_{(i_1 \cdots i_m)} \sum_{(j_1 \cdots j_m)} \sigma^{i_1 \cdots i_m, j_1 \cdots j_m} \gamma_{i_1 \cdots i_m} \gamma_{j_1 \cdots j_m} \tag{7.4}$$

where

$$\sigma^{i_1 \cdots i_m, j_1 \cdots j_m} = \frac{\text{Cofactor of } E\{\gamma_{i_1 \cdots i_m} \cdot \gamma_{j_1 \cdots j_m}\} \text{ in } A_m}{|A_m|}$$

and each summation is over the $(\theta_m - \tau_m)$ sets of $m$ integers $(i_1, i_2, \cdots, i_m)$ which appear as subscripts of the $\gamma_{i_1 \cdots i_m}$.

$$\begin{aligned} E\{Q_m\} &= \sum_{(i_1 \cdots i_m)} \frac{1}{|A_m|} \sum_{(j_1 \cdots j_m)} [\text{Cofactor of } E\{\gamma_{i_1 \cdots i_m} \cdot \gamma_{j_1 \cdots j_m}\}] \\ &\qquad \cdot E\{\gamma_{i_1 \cdots i_m} \cdot \gamma_{j_1 \cdots j_m}\} \\ &= \sum_{(i_1 \cdots i_m)} \cdot \frac{|A_m|}{|A_m|} = \theta_m - \tau_m. \end{aligned} \tag{7.5}$$

**8. The Case $m = k$.** Let $h$ index the permutations of $(1, 2, \cdots, k)$. Let $n_h$ be the number of times that ordering, $h$, occurs in a sample of $n$, and let $p_h$ be the probability of that ordering.

Then

$$\gamma_h = \frac{(n_h - n \cdot p_h)}{\sqrt{np_h(1 - p_h)}} \tag{8.1}$$

and

$$\sum_{h=1}^{k!} (1 - p_h)\gamma_h^2 = \sum_h \frac{(n_h - n \cdot p_h)^2}{n \cdot p_h}$$

which is immediately recognizable as Pearson's Chi-square statistic.

THEOREM VI: $Q_k = \sum_{h=1}^{k!} (1 - p_h)\gamma_h^2$.

PROOF: It has been shown that $Q_k$ is asymptotically Chi-square with $(k! - 1)$ degrees of freedom, and it is well-known that Pearson's statistic has the same limiting distribution. Regarding one of the $\gamma_h$ as a linear combination of the remainder, both statistics are quadratic forms of rank $(k! - 1)$ in the same $(k! - 1)$ variates with the same asymptotic Chi-square distribution. This is exactly the situation considered in Lemma 1, (Section 6), and by an identical argument, the two statistics are equal.

$$\therefore Q_k = \sum_h (1 - p_h)\gamma_h^2.$$

**9. Consistency and the classification of alternatives.** To make a symmetry test of order $m$, we compute $Q_m$ and reject the null hypothesis if $Q_m$ is large. But, in a particular case, what order should the test be? The answer to this question requires a consideration of the alternative hypothesis. Intuitively, the tests of low order such as Friedman's $X_r^2$, provide a relatively high sensitivity to a small class of alternatives, whereas the high order tests give a low sensitivity—thus

requiring a large sample—against a large class of alternatives. This concept of a classification of alternatives is made more precise by the following definition and Theorem VII.

*Definition:* An alternative is *distinct* from the null hypothesis *at level m* if, under the alternative

$$E\{Z_{i_1 i_2 \cdots i_m}\} \neq p_{i_1 i_2 \cdots i_m}$$

for at least one choice and permutation of the digits $(i_1, i_2, \cdots, i_m)$.

*Remark:* Since $Z_{i_1 i_2 \cdots i_m}$ can be written as a sum of such counter variables of higher order, it is immediate that, if an alternative is distinct at level $m$, it is distinct at any level $m' > m$.

THEOREM VII: *The symmetry test of order m is consistent against any alternative which is distinct at level m, and only against such alternatives.*

PROOF: The null hypothesis is rejected whenever $Q_m > c$, i.e., whenever the point $\gamma_m = \{\gamma_{i_1 \cdots i_m}\}$ lies outside the region, $S$, defined by

$$S = [\gamma_m \mid Q_m \leqq c]. \tag{9.1}$$

$c$ is chosen so that the measure of $S$, computed with the asymptotic distribution under the null hypothesis is $1 - \alpha$. $S$ is then a fixed, finite region. Under any hypothesis the measure of $S$ tends to a specific value, $P$, as $n$ approaches infinity: we wish to show (1) that this value is zero under alternatives distinct at level $m$, and (2) that this value is different from zero under alternatives not distinct at level $m$.

The measure of $S$ can be written in the form $P + \eta$, where $P$ is the measure of $S$ under the approximating normal distribution with the same mean and variance-covariance matrix as the given distribution, and $\eta$ is a correction term which approaches zero as $n$ approaches infinity. The variances and covariances of $\gamma_m$ are independent of $n$; and the mean of $\gamma_m$, and of the approximating normal variate, has coordinates

$$E\{\gamma_{i_1 \cdots i_m}\} = \sqrt{\frac{n}{p_{i_1 \cdots i_m}(1 - p_{i_1 \cdots i_m})}}\,[E\{Z_{i_1 \cdots i_m}\} - p_{i_1 \cdots i_m}] \tag{9.2}$$

Under any alternative not distinct at level $m$, these coordinates are all zero: the approximating normal distribution, and therefore $P$, is independent of $n$. Clearly, $P \neq 0$, thus establishing the second part of the theorem.

Under an alternative distinct at level $m$, however, the distance, $d$, from the mean of the distribution to the origin given by

$$d^2 = n \sum_{(i_1 \cdots i_m)} \frac{[E\{Z_{i_1 \cdots i_m}\} - p_{i_1 \cdots i_m}]^2}{p_{i_1 \cdots i_m}(1 - p_{i_1 \cdots i_m})} \neq 0. \tag{9.3}$$

Thus the distance from the mean to the origin approaches infinity as $n$ tends to infinity.

It is well-known that the ordinate of a normal distribution tends to zero as the distance from the mean increases; hence, the measure, $P$, of the fixed, bounded

set $S$, under the sequence of approximating normal distributions, tends to zero. Since both $P$ and $\eta$ tend to zero under an alternative distinct at level $m$, as $n$ approaches infinity, the consistency of the test against such alternatives is established.

*Acknowledgements.* The author wishes to express his gratitude to Professor S. S. Wilks for his guidance and encouragement. He is also indebted to Professor J. W. Tukey for suggesting the original problem that led to this study; and to Professor D. A. S. Fraser of Toronto, for many stimulating suggestions.

REFERENCES

[1] MILTON FRIEDMAN, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Stat. Assn.*, Vol. 32 (1937), pp. 675–701.

[2] T. J. TERPSTRA, "A non-parametric test for the problem of $k$ samples," *Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, Vol. 57 (1954), pp. 505–512.

[3] W. J. DIXON, "Power functions of the sign test and power efficiency for normal alternatives," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 467–473.

[4] H. CRAMER, "Random Variables and Probability Distributions," *Cambridge Tracts in Mathematics, No.* 36, Cambridge, 1937.

[5] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1947.

[6] H. CRAMER, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

# CONTRIBUTIONS TO THE THEORY OF RANK ORDER STATISTICS—THE ONE-SAMPLE CASE[1]

I. Richard Savage

*University of Minnesota*

**0. Summary.** The one-sample problem is considered using techniques developed earlier [2], [3]. Let $Z = (Z_1, \cdots, Z_N)$ be a random vector with $Z_i = 1(0)$ if the $i$th smallest in absolute value in a sample of $N$ from the density $f(x)$ is positive (negative). Then

$$P(Z = z) = N! \int_{0 \le y_1 \le \cdots \le y_N \le \infty} \cdots \int \prod_{i=1}^{N} [f^{1-z_i}(-y_i) f^{z_i}(y_i)\, dy_i]$$

Conditions are found implying $P(Z = z) > P(Z = z')$ where $z$ is derived from $z'$ by replacing a 0 by a 1, or interchanging a 0 and 1 in $z'$ by moving the 1 to the right. These conditions are met by the normal and other distributions.

The results are useful in finding good tests of such null hypotheses as $X_1, \cdots, X_N$ are independently and identically distributed symmetrically about zero against such alternatives as slippage to the right. The Wilcoxon one sample signed rank test is a typical nonparametric procedure used under these conditions [4].

**1. Assumptions and notations.** Throughout it is assumed that $X_1, \cdots, X_N$ are independently and identically distributed random variables with a continuous distribution function, $F(x, \theta)$ having a density function $f(x, \theta)$. $\theta$ will be a real valued parameter and under the *null hypothesis* $H_0: \theta = 0$.

If $x_1, \cdots, x_N$ are the observations and $y_1, \cdots, y_N$ are the absolute values of the observations arranged from smallest to largest, then $z = (z_1, \cdots, z_N)$ is defined to be the observed *rank order* where $z_i = 1$ if $y_i$ is the absolute value of a positive number and $z_i = 0$ if $y_i$ is the absolute value of a negative number. Thus, $n = \sum_{i=1}^{N} z_i$ is the number of positive observations and $m = \sum_{i=1}^{N}(1 - z_i)$ is the number of negative observations. Corresponding to the observed $y = (y_1, \cdots, y_N)$ and $z = (z_1, \cdots, z_N)$ are the random variables $Y = (Y_1, \cdots, Y_N)$ and $Z = (Z_1, \cdots, Z_N)$. There are $2^N$ possible values of $Z$. For a specified value of $n$ there are $\binom{N}{n}$ values of $Z$. For $n$ fixed the conditional distribution of $Z$ is that of the two sample problem [2] where the first population has the c.d.f.

$$F^{-}(x, \theta) = \begin{cases} \dfrac{F(0, \theta) - F(-x, \theta)}{F(0, \theta)}, & x \ge 0, \\ 0, & x < 0, \end{cases}$$

Received June 23, 1958; revised January 22, 1959.

[1] Work supported in part by the Office of Naval Research.

and the second population has c.d.f.

$$F^{+}(x,\theta) = \begin{cases} \dfrac{F(x,\theta) - F(0,\theta)}{1 - F(0,\theta)}, & x \geqq 0, \\ 0, & x < 0. \end{cases}$$

Thus for fixed $n$ the partial order problem is exactly that treated in [2] where $F(x) = F^{-}(x)$ and $G(x) = F^{+}(x)$. The previous results are not immediately applicable, however, since it is not clear how to impose conditions on $F(x, \theta)$ in order to get $F^{-}(x, \theta)$ and $F^{+}(x, \theta)$ to satisfy the conditions of [2]. In Section 2, the case of fixed $n$ is considered. The notation $z'Lz$ denotes the following relationship: $z_k' = z_k$ for all $k = 1, \cdots, N$ except $i$ and $j$ $(i < j)$ and $z_i = z_j' = 0$, $z_j = z_i' = 1$. This notation is also used if there exists $z^1, \cdots, z^I$ such that $z'Lz^1 \cdots z^ILz$, e.g., $(1010)L(0101)$ since $(1010)L(0110)$ and $(0110)L(0101)$.

In Section 3, the partial order of the probabilities of two rank orders having different values of $n$ is considered. The notation $z'Sz$ denotes $z_k \geqq z_k'$ for $k = 1, \cdots, N$ and $>$ holds for at least one value of $k$.

The following formula is used repeatedly:

$$(1.1)\quad P(Z = z) = N! \int_{0 \leq y_1 \leq \cdots \leq y_N \leq \infty} \cdots \int \prod_{i=1}^{N} [f^{z_i}(y_i, \theta) f^{1-z_i}(-y_i, \theta)\, dy_i]$$

The null hypothesis of concern is $F(-x, 0) + F(x, 0) = 1$, i.e., symmetry about 0. Under $H_0$, $P(Z = z) = 2^{-N}$ for each $z$. An alternative of particular interest is

$$F(x,\theta) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-(t-\theta)^2/2}\, dt, \qquad \theta > 0.$$

All of the following results apply to this alternative hypothesis.

## 2. The case of fixed $n$.

THEOREM 2.1:

a) $f(x, \theta) = u(x)v(\theta)e^{a(x)b(\theta)}$

b) $v(\theta) \geqq 0$

c) $u(x) = u(-x) > 0$

d) If $x < y$ then $a(x) < a(y)$

e) $b(\theta) > 0$

*then* $z'Lz$ *implies* $\Delta = P(Z = z) - P(Z = z') > 0$.

PROOF: Using (1.1) obtain

$$\Delta = N! \int_{0 \leq y_1 \leq \cdots \leq y_N \leq \infty} \cdots \int A(y_i, y_j) \prod_{i=1}^{N} [f^{z_i}(y_i, \theta) f^{1-z_i}(-y_i, \theta)\, dy_i]$$

where

$$A(y_i, y_j) = 1 - \frac{f(y_i, \theta)f(-y_j, \theta)}{f(-y_i, \theta)f(y_j, \theta)}$$
$$= 1 - \exp \{b(\theta)[a(y_i) - a(-y_i) + a(-y_j) - a(y_j)]\}$$

The theorem is proved by showing $A(y_i, y_j) \geqq 0$, which follows since the exponent is negative due to the monotonicity of $a(x)$.

THEOREM 2.2: *If*

a) $f(x, \theta) = f(x - \theta) = f(\theta - x)$

b) *If* $x > y$ *and* $\theta > \delta$ *then,*

$$\begin{vmatrix} f(x, \theta) & f(x, \delta) \\ f(y, \theta) & f(y, \delta) \end{vmatrix} > 0$$

c) $\theta > 0$

*then* $z'Lz$ *implies* $\Delta = P(Z = z) - P(Z = z') > 0$.

PROOF:

$$\Delta = N! \int_{0 \leq y_1 \leq \cdots \leq y_N \leq \infty} \cdots \int B(y_i, y_j) \left\{ \prod_{\substack{k=1 \\ i \neq k \neq j}}^{N} [f^{z_k}(y_k, \theta) f^{1-z_k}(-y_k, \theta)] \right\} \left[ \prod_{k=1}^{N} dy_k \right]$$

where $B(y_i, y_j) = f(y_j, \theta)f(-y_i, \theta) - f(-y_j, \theta)f(y_i, \theta)$ and the proof is completed by showing $B(y_i, y_j) > 0$. In assumption b let $x = y_j$, $y = y_i$, and $\delta = -\theta$ so that

$$0 < \begin{vmatrix} f(y_j, \theta) & f(y_j, -\theta) \\ f(y_i, \theta) & f(y_i, -\theta) \end{vmatrix} = f(y_j - \theta)f(y_i + \theta) - f(y_j + \theta)f(y_i - \theta).$$

Now use $f(x - \theta) = f(\theta - x)$, assumption a, hence

$$0 < f(y_j - \theta)f(y_i + \theta) - f(y_j + \theta)f(y_i - \theta)$$
$$= f(y_j, \theta)f(-y_i, \theta) - f(-y_j, \theta)f(y_i, \theta)$$
$$= B(y_i, y_j).$$

## 3. The case of variable $n$.

THEOREM 3.1: *Under the assumptions of Theorem* 2.1, *if* $z'Sz$, *then* $\Delta = P(Z = z) - P(Z = z') > 0$.

PROOF. It is sufficient to consider only the special case $z_k' = z_k$ for all $k = 1, \cdots, N$ except $k = i$ where $z_i = 1$ and $z_i' = 0$. Then,

$$\Delta = N! \int_{0 \leq y_1 \cdots \leq y_N \leq \infty} \cdots \int C(y_i) \left\{ \prod_{k=1}^{N} [f^{z_k}(y_k, \theta) f^{1-z_k}(-y_k, \theta)\, dy_k] \right\}$$

and the proof is completed by showing $C(y_i) = 1 - f(-y_i, \theta) \times [f(y_i, \theta)]^{-1} > 0$. Using the special form of $f(x, \theta)$, $C(y_i) = 1 - \exp\{b(\theta)[a(-y_i) - a(y_i)]\}$ and again the exponent is negative because of the monotonicity of $a(y)$.

THEOREM 3.2: *If*

a) $f(x, \theta) = f_\theta(x - \theta) = f_\theta(\theta - x)$

b) *If* $x > y > 0$ *then* $f_\theta(y) > f_\theta(x)$

c) $\theta > 0$

*then* $z' S z$ *implies* $\Delta = P(Z = z) - P(Z = z') > 0$.

PROOF:

$$\Delta = N! \int_{0 \le y_1 \le \cdots \le y_N \le \infty} \cdots \int D(y_i) \left\{ \prod_{\substack{k=1 \\ i \ne k}}^{N} [f^{z_k}(y_k, \theta) f^{1-z_k}(-y_k, \theta)] \right\} \left[ \prod_{k=1}^{N} dy_k \right]$$

and it is sufficient to show that $D(y_i) = f(y_i, \theta) - f(-y_i, \theta) > 0$. First, using assumption a, $D(y_i) = f_\theta(y_i - \theta) - f_\theta(-y_i - \theta) = f_\theta(y_i - \theta) - f_\theta(y_i + \theta)$. Now if $y_i > \theta$ the result follows from b, since $y_i - \theta < y_i + \theta$. If $y_i < \theta$ the result follows from b when we write $D(y_i) = f_\theta(\theta - y_i) - f_\theta(y_i + \theta)$.

*Remark* 1. In Theorem 3.2 writing $f(x, \theta) = f_\theta(x - \theta)$ allows $f(x, \theta)$ not only to be translations of the $H_0$ but also other changes, such as changes in scale, can occur.

*Remark* 2. The assumptions of Theorem 2.2 imply those of Theorem 3.2 but not conversely. If in b of Theorem 2.2 we set $\delta = 0$, $2\theta = x + y$, and $0 < y < x$ we obtain b of Theorem 3.2. The Cauchy density is a counter example of the converse.

**4. Some partial orderings.** If the assumptions of Theorems 2.1 and/or of 2.2 and 3.2 hold, then the following diagrams are obtained:

$N = 1$

$$1 \to 0$$

(where $P(Z = z) > P(Z = z') \equiv z \to z'$)

$N = 2$

$$11 \to 01 \to 10 \to 00$$

$N = 3$

$$\begin{tikzcd}
111 \arrow[r] & 011 \arrow[r] & 101 \arrow[r] \arrow[dr] & 110 \arrow[dr] & & & \\
& & & 001 \arrow[r] & 010 \arrow[r] & 100 \arrow[r] & 000
\end{tikzcd}$$

$N = 4$

$$\begin{tikzcd}
1111 \arrow[r] & 0111 \arrow[r] & 1011 \arrow[r] \arrow[dd] & 1101 \arrow[r] \arrow[dd] & 1110 \arrow[d] & & & & \\
& & & & 0110 \arrow[dr] & & & & \\
& & 0011 \arrow[r] & 0101 \arrow[ur] \arrow[dr] & & 1010 \arrow[r] \arrow[dd] & 1100 \arrow[dd] & & \\
& & & & 1001 \arrow[ur] \arrow[d] & & & & \\
& & & & 0001 \arrow[r] & 0010 \arrow[r] & 0100 \arrow[r] & 1000 \arrow[r] & 0000
\end{tikzcd}$$

Now consider the uniform distribution $f(x, \theta) = 1$ for $\theta - \frac{1}{2} \leqq x \leqq \theta + \frac{1}{2}$ and 0 otherwise, $0 \leqq \theta \leqq \frac{1}{2}$. If $n'$ is the length of the last run of 1's in $z$ or $n'$ = the number of the positive observations greater than the maximum of the absolute values of the negative observations, then

$$P(Z = z) = \sum_{i=0}^{n'} \binom{N}{i} (\tfrac{1}{2} - \theta)^{N-i}(2\theta)^i \tag{4.1}$$

To obtain (4.1), begin with

$$P(Z = z) = \sum_{i=0}^{n'} P(Z = z \mid i \text{ observations} > \tfrac{1}{2} - \theta) \times P(i \text{ observations} > \tfrac{1}{2} - \theta)$$

and use

$$P(Z = z \mid i \quad \text{observations} \quad > \tfrac{1}{2} - \theta) = 2^{-(N-i)},$$

$$P(i \quad \text{observations} \quad > \tfrac{1}{2} - \theta) = \binom{N}{i} (2\theta)^i(1 - 2\theta)^{N-i}.$$

Holding $\theta$ fixed, $P(Z = z)$ is an increasing function of $n'$, and otherwise does not depend on $z$. Thus, the most powerful rank order tests depend solely on $n'$.

**5. A statistical application.** For the normal alternative hypotheses, mentioned at the end of Section 1, several test statistics have been proposed:

a. On intuitive grounds Wilcoxon proposed the statistic

$$T_W = \sum_{i=1}^{N} z_i i.$$

b. Fraser [1] showed the locally most powerful rank order test is of the form $T_F = \sum_{i=1}^{N} z_i E(X_{Ni})$ where $X_{Ni}$ is the $i$th order statistic from the chi distribution with one degree of freedom.

Both of these statistics are of the form $T = \sum_{i=1}^{N} z_i a_i$ where the $a_i$ form an increasing sequence. It is easily verified that if $z'Lz$ and/or $z'Sz$ then $T(z) > T(z')$. Thus statistics of this form take full advantage of the results of this paper, i.e., using these statistics the known more probable rank orders are put into the critical region first.

**6. Normal slippage.** The theorems of Sections 2 and 3 do not help in the ordering of $P_1 = P(Z = (0, 0, 1))$ and $P_2 = P(Z = (1, 1, 0))$, for normal alternatives. If $P_1 > P_2$ then the partial order for $N = 3$ given in Section 4 becomes the simple order:

$$111 \to 011 \to 101 \to 001 \to 110 \to 010 \to 100 \to 000$$

THEOREM 6.1:[2] *If $X_1, \cdots, X_N (N \geqq 3)$ are independently and normally distributed, each with mean $\theta (>0)$ and variance* 1, *then* $\Delta = P(Z = z) -$

[2] M. Sobel proved this result for $N = 3$ at the 1958 Summer Statistical Institute sponsored by the National Science Foundation.

*$P(Z = z') > 0$ where $z$ and $z'$ are identical except $z_1 = z_2 = z_3' = 0$ and $z_1' = z_2' = z_3 = 1$.*

PROOF: Using (1.1)

$$\Delta = \frac{N!}{(2\pi)^{3/2}} \int \cdot \quad \cdot \quad \cdot \int \left\{ \prod_{i=4}^{N} [f^{z_i}(y_i, \theta) f^{1-z_i}(-y_i, \theta)] \right\} \left[ \prod_{i=1}^{N} dy_i \right]$$
$$\times \{\exp [-\tfrac{1}{2} (y_1^2 + y_2^2 + y_3^2 + 3\theta^2)]\} \times [e^{\theta(y_3 - y_1 - y_2)} - e^{\theta(y_1 + y_2 - y_3)}]$$

Now make the transformation $y_1 = w_1$, $y_2 = w_1 + w_2$, $y_3 = w_1 + w_2 + w_3$, and $y_i = w_i$ for $i = 4, \cdots, N$. The Jacobian is 1 and the region of integration becomes $0 \leqq w_i \leqq w_4 \leqq w_5 \leqq \cdots \leqq w_N \leqq \infty$ for $i = 1, 2, 3$; and $\sum_{i=1}^{3} w_i \leqq w_4$. Then

$$\Delta = \frac{N!}{(2\pi)^{3/2}} \int \cdot \quad \cdot \quad \cdot \int \left\{ \prod_{i=4}^{N} [f^{z_i}(w_i, \theta) f^{1-z_i}(-w_i, \theta)] \right\} \left[ \sum_{i=1}^{N} dw_i \right]$$
$$\times \{\exp [-\tfrac{1}{2} (w_1^2 + (w_1 + w_2)^2 + (w_1 + w_2 + w_3)^2 + 3\theta^2)]\}$$
$$\times [e^{(\theta w_3 - w_1)} - e^{\theta(w_1 - w_3)}]$$

The above integral is equivalent to the following integral, where the region of integration is $0 \leqq w_1 \leqq w_3 \leqq w_4 \cdots \leqq w_N \leqq \infty$, $0 \leqq w_2 \leqq w_4$, and $\sum_{i=1}^{3} w_i \leqq w_4$.

$$\Delta = \frac{N!}{(2\pi)^{3/2}} \int \cdot \quad \cdot \quad \cdot \int \left[ \prod_{i=4}^{N} [f^{z_i}(w_i, \theta) f^{1-z_i}(-w_i, \theta)] \right] \left[ \prod_{i=1}^{N} dw_i \right]$$
$$\times \{\exp [-\tfrac{1}{2} (w_1^2 + (w_1 + w_2)^2 + (w_1 + w_2 + w_3)^2 + 3\theta^2)]$$
$$- \exp [-\tfrac{1}{2} (w_3^2 + (w_3 + w_2)^2 + (w_3 + w_2 + w_1)^2 + 3\theta^2)]\}$$
$$\times [e^{\theta(w_3 - w_1)} - e^{\theta(w_1 - w_3)}]$$

For the region of integration each of the factors in the above integrand is clearly $>0$ except for the $\{\quad\}$. To show $\{\quad\} > 0$, prove the equivalent inequality

$$w_3^2 + (w_3 + w_2)^2 > w_1^2 + (w_1 + w_2)^2 \equiv w_3(w_3 + w_2) > w_1(w_1 + w_2)$$

which is clearly the case since $w_3 > w_1 > 0$.

Theorem 6.1 implies a simple order for the five most probable rank orders against normal slippage.

## REFERENCES


[1] D. A. S. FRASER, "Most powerful rank-type tests", *Ann. Math. Stat.*, Vol. 28 (1957), pp. 1040–1043.

[2] I. R. SAVAGE, "Contributions to the theory of rank order statistics—the two sample cases", *Ann. Math. Stat.*, Vol. 27, (1956), pp. 590–615.

[3] I. R. SAVAGE, "Contributions to the theory of rank order statistics—the 'trend case' ", *Ann. Math. Stat.*, Vol. 28, (1957), pp. 968–977.

[4] F. WILCOXON, *Some Rapid Approximate Statistical Procedures*, American Cyanamid Co., Statistical Research Laboratories (July, 1949), p. 16.

# THE DISTRIBUTION OF A GENERALIZED $D_n^+$ STATISTIC[1]

By Meyer Dwass

*Northwestern University*

**1. Introduction and summary.** Let $F_n(x)$ be the empirical c.d.f. of $n$ independent random variables, each distributed according to the same continuous c.d.f. $F(x)$. The major object of this paper is to obtain in explicit form the probability law of the random variable

$$D_n^+(\gamma) = \sup_{-\infty < x < \infty} \{F_n(x) - \gamma F(x)\}.$$

It is no loss of generality to suppose that $F(x)$ is the c.d.f. of the uniform distribution on [0, 1], so this assumption will be held throughout the paper.

When $\gamma = 1$, then $D_n^+(1)$ is the usual one-sided goodness of fit statistic whose asymptotic distribution was first derived by Smirnov [6]. We obtain in several different forms (formulas 2.2 and 2.3) an expression for

$$P(D_n^+(\gamma) < a) = P(F_n(x) \leqq a + \gamma x, 0 \leqq x \leqq 1).$$

Formula (2.2) agrees with the one found by Birnbaum and Tingey [2] when $\gamma = 1$, which is the "classical" case. As a matter of fact, it seems to have been overlooked that this formula, for finite $n$, had already appeared in a paper by Smirnov [6]. The new formula (2.3) would seem to involve fewer computations for actual numerical evaluation. One rather remarkable fact which results from (2.3) is that

$$P(F_n(x) \leqq \gamma x, 0 \leqq x \leqq 1) = \begin{cases} 1 - \dfrac{1}{\gamma}, & \gamma > 1 \\ 0, & \gamma \leqq 1, \end{cases}$$

*for any $n$*. This was noted by Daniels [4] and was rediscovered by Robbins [5].

Using (2.3) it is easy to evaluate $\lim_{n\to\infty} P(F_n(x) \leqq a(n) + \gamma x)$ where $\gamma$, $(\gamma > 1)$ is fixed and $a(n) = d/n$, where $d$ is fixed. The limiting distribution when $\gamma > 1$ can be used to derive some facts about the Poisson Process which were recently discovered by Baxter and Donsker [1].

The methods used are elementary. To assist the reader, the results are all listed in Section 2 and Section 3 is devoted to giving proofs.

**2. Statement of results.** First a few pieces of notation are introduced. Let

$$P_n(a, \gamma) = P(F_n(x) < a + \gamma x,\ 0 \leqq x \leqq 1) = P(D_n^+(\gamma) < a),$$

---

Received January 20, 1959; revised April 1, 1959.

[1] Sponsored by Air Force Office of Scientific Research Air Research and Development Command, Contract No. AF 49 (638)-151.

and let

$$C_n(a, \gamma, i) = \binom{n}{i}\left(\frac{n-i}{n\gamma} - \frac{a}{\gamma}\right)^{n-i}\left(1 - \left(\frac{n-i}{n\gamma} - \frac{a}{\gamma}\right)\right)^{i-1}\left(\frac{\gamma + a - 1}{\gamma}\right).$$

For simplicity, whenever it is reasonable to do so, $P_n$ and $C_n(i)$ are used instead of the more complicated symbols.

It is assumed hereafter that

$$0 < a < 1, \qquad a + \gamma > 1, \qquad \gamma > 0, \tag{2.1}$$

for otherwise $P_n$ becomes trivially either 0 or 1.

THEOREM 1:

$$P_n = 1 - \sum_{i=0}^{k} C_n(i), \tag{2.2}$$

*or, equivalently,*

$$P_n = \sum_{i=k+1}^{n} C_n(i), \tag{2.3}$$

*where the integer k is defined by*

$$\frac{k}{n} \leqq (1 - a) < \frac{k+1}{n}.$$

*Remark on Theorem* 1: When $\gamma = 1$, formula (2.2) agrees with the result of Birnbaum and Tingey. However, when $a$ is of the order of $1/\sqrt{n}$, (2.3) will usually require many fewer values of $C_n(i)$ to compute. For example, for $n = 50$, Table 1 of [2] indicates that $a$ varies roughly between $\frac{1}{7}$ and $\frac{1}{4}$, for those probabilities "interesting" for statistical applications. Hence the number of $C_n(i)$ terms to be computed using (2.3) ranges from about 37 to 42 for these $a$'s, whereas using (2.2) the range is from 7 to 12 terms.

Setting $a = 0$ in (2.3) yields a

COROLLARY TO THEOREM 1 (Daniels [4], Robbins [5]):

$$P(F_n(x) < \gamma x, \quad 0 \leqq x \leqq 1) = \begin{cases} 1 - \dfrac{1}{\gamma}, & \gamma > 1, \\ 0, & \gamma \leqq 1. \end{cases}$$

It is interesting that this result does not depend on $n$.

THEOREM 2: *Let $a = d/n$, where $d$ is a fixed positive real number, and let $\gamma$ be greater than* 1. *Then*

$$\lim_{n\to\infty} P_n(d/n, \gamma) = \left(1 - \frac{1}{\gamma}\right)\sum_{i=0}^{[d]} \frac{1}{i!}\left(\frac{i-d}{\gamma}\right)^i e^{(d-i)/\gamma}. \tag{2.4}$$

*Remarks on Theorem* 2:

a) The interesting fact here is that when $\gamma > 1$ the proper norming for $a$ requires it to be of the order of $1/n$ rather than $1/\sqrt{n}$ as in the case $\gamma = 1$.

Contrary to what one would expect, the derivation is much more elementary when $\gamma > 1$ than when $\gamma = 1$.

b) The right hand side of (2.4) is the same as an expression obtained by Baxter and Donsker [1] in connection with the Poisson process. Theorem 2 immediately gives the same result which is summarized in the following corollary.

Corollary to Theorem 2: *Let* $Y(t), 0 \leqq t < \infty$ *be the Poisson process with stationary and independent increments and parameter* $\lambda > 0$, *and* $Y(0) = 0$. *Let* $\gamma > \lambda$, *and* $d$ *be positive. Then*

$$P(Y(t) < d + \gamma t, 0 \leqq t < \infty) = \left(1 - \frac{\lambda}{\gamma}\right) \sum_{i=0}^{[d]} \frac{1}{i!} \left(\frac{\lambda}{\gamma}\right)^i (i - d)^i e^{(\lambda/\gamma)(d-i)}.$$

**3. Proofs.**

I. *Proof of Theorem* 1, *equation* (2.2). The basic idea used in the proof is the following: let $x_1 < x_2 < \cdots < x_n$ be the ordered values of $n$ independent random variables, each uniformly distributed over (0, 1). Then, it is well known that given $x_n$, the conditional distribution of

$$x_1/x_n, \cdots, x_{n-1}/x_n$$

is that of the ordered values of $(n - 1)$ independent random variables, each uniformly distributed over (0, 1). Using this fact it is easy to verify the following conditional probability statements:

$$P(F_n(x) < a + \gamma x \mid x_n = t)$$

$$= \begin{cases} 1, & \text{if } \dfrac{1-a}{\gamma} < t \text{ and } \dfrac{n-1}{n} \leqq a \leqq 1, \\ P_{n-1}\left(\dfrac{n}{n-1} a, \dfrac{n}{n-1} \gamma t\right), & \text{if } \dfrac{1-a}{\gamma} < t \text{ and } a < \dfrac{n-1}{n}, \\ 0, & \text{if } t \leqq \dfrac{1-a}{\gamma}. \end{cases}$$

Using the fact that the frequency function of $x_n$ is

$$\begin{aligned} &nt^{n-1}, \quad 0 \leqq t \leqq 1, \\ &0, \quad \text{otherwise}, \end{aligned}$$

we have the basic recursion relationship

$$(3.1) \qquad P_n(a, \gamma) = \begin{cases} \displaystyle\int_{\frac{1-a}{\gamma}}^1 P_{n-1}\left(\frac{n}{n-1} a, \frac{n}{n-1} \gamma t\right) nt^{n-1}\, dt, & \text{if } a < \dfrac{n-1}{n} \\ \displaystyle\int_{\frac{1-a}{\gamma}}^1 nt^{n-1}\, dt = 1 - \left(\frac{1-a}{\gamma}\right)^n, & \text{if } \dfrac{n-1}{n} \leqq a \leqq 1. \end{cases}$$

An induction argument can now be applied to prove (2.2). Its truth is trivially

true when $n = 1$. Assume now that it holds for arbitrary $n$. By this induction hypothesis,

$$P_n\left(\frac{n+1}{n}a, \frac{n+1}{n}\gamma t\right) = 1 - \sum_{i=0}^{k} C_n\left(\frac{n+1}{n}a, \frac{n+1}{n}\gamma t, i\right),$$

where $k$ is defined by $k/n \leqq 1 - ((n+1)/n)\,a < (k+1)/n$, or equivalently by $(k+1)/(n+1) \leqq (1-a) < (k+2)/(n+1)$. By a routine, but tedious, computation which is omitted

$$\int_{\frac{1-a}{\gamma}}^{1} C_n \frac{n+1}{n}a, \frac{n+1}{n}\gamma t, i)(n+1)\,t^n\,dt = C_{n+1}(a, \gamma, i+1)\,.$$

Hence, applying (3.1), it follows that (2.2) is true for $n+1$, which completes the proof of the first part of Theorem 1.

II. *Proof of Theorem* 1, *equation* (2.3). This follows from 2.2 by means of part a) of the following lemma:

LEMMA:

a) $\displaystyle\sum_{i=0}^{n} C_n(i) = 1$

b) $\displaystyle\sum_{i=0}^{n} \binom{n}{i} (A+i)^i(B-i)^{n-i-1} = (A+B)^n/(B-n)$

c) $\displaystyle\sum_{i=0}^{n} \frac{1}{i+1}\binom{n}{i} (A+i)^i(B-i)^{n-i-1} = \frac{1}{(A-1)(B-n)(n+1)}$

$$\cdot\,[(A+B)^n(A+B-n-1) - (B+1)^n(B-n)]$$

d) $\displaystyle\sum_{j=0}^{n-1} \binom{n}{j+1} (A+j)^j(B-j)^{n-j-1} = \frac{(A+B)^n - (B+1)^n}{A-1}$,

where $A \neq 1$, $B \neq n$. Part b) is a formula of Abel's which is referred to in Lemma 1 of [3]. Part c) is proved in [3]. Part d) is proved by writing

$$\sum_{j=0}^{n-1} \binom{n}{j+1} (A+j)^j(B-j)^{n-j-1}$$

$$= (n+1)\sum_{j=0}^{n} \frac{1}{j+1}\binom{n}{j}(A+j)^j(B-j)^{n-j-1} - \sum_{j=0}^{n}\binom{n}{j}(A+j)^j(B-j)^{n-j-1},$$

and by then applying b) and c). Part a) now follows from d) as follows. $C_n(i)$ can be expressed as

$$C_n(i) = \frac{1}{(n\gamma)^n}(n - na - 1 - (i-1))^{n-1-(i-1)}$$

$$.\,(n\gamma - n + na + 1 + (i-1))^{i-1}\left(\frac{\gamma + a - 1}{\gamma}\right)\binom{n}{(i-1)+1}.$$

Now let $i - 1 = j$, $n\gamma - n + na + 1 = A$, $n - na - 1 = B$, then

$$\sum_{i=0}^{n} C_n(i) = \frac{1}{(n\gamma)^n}\left(\frac{\gamma + a - 1}{\gamma}\sum_{j=1}^{n-1}\binom{n}{j+1}(A + j)^j(B - j)^{n-1-j}\right.$$

and a) follows from d) by routine algebra. The lemma is completely proved. Now part a) of the lemma immediately implies (3.2), given the truth of (3.1).

III. *Proof of Theorem* 2. This follows in a routine way from (2.3).

IV. *Proof of corollary to Theorem* 2. It is sufficient to suppose that $\lambda = 1$, the general case easily following from this special one.

Let $A(T)$ be the event that

$$\left\{\frac{Y(t)}{Y(T)} \leq \frac{d}{Y(T)} + \gamma\frac{t}{T}, \qquad 0 \leq t \leq T\right\},$$

where $Y(t)/Y(T)$ can be defined as 0 if $Y(T) = 0$.

According to the well-known relationship between the Poisson process and uniformly distributed random variables,

$$P(A(T) \mid Y(T) = n) = P_n(d/n, \gamma), \qquad n \geq 1.$$

Hence

$$P(A(T)) = \sum_{n=0}^{\infty} P(A(T) \mid Y(T) = n)\frac{e^{-T}T^n}{n!} = \sum_{n=0}^{\infty} P_n(d/n, \gamma)\frac{e^{-T}T^n}{n!}.$$

Since $P_n(d/n, \gamma)$ approaches the right side of (2.4) as $n \to \infty$, and since $\sum_{n=0}^{r} e^{-T}T^n/n! \to 0$ as $n \to \infty$ for any fixed $r$, an easy argument proves that

$$P(A(T)) \to \lim_{n\to\infty} P_n(d/n, \gamma) \quad \text{as} \quad T \to \infty.$$

Since $Y(T)/T$ converges to 1 with probability 1, it is not hard to show that

$$\lim_{T\to\infty} P(A(T)) = P(Y(t) \leq d + \gamma t, 0 \leq t < \infty),$$

which completes the proof.

## REFERENCES


[1] Glen Baxter and M. D. Donsker, "On the distribution of the supremum functional for processes with stationary independent increments," *Trans. Am. Math. Soc.*, Vol. 85 (1957), pp. 73–87.

[2] Z. W. Birnbaum and F. H. Tingey, "One sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[3] Z. W. Birnbaum and R. Pyke, "On some distributions related to the statistic $D_n^+$," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 179–187.

[4] H. E. Daniels, "The statistical theory of the strength of bundles of threads. I," *Proc. Roy. Soc., Ser. A*, Vol. 183 (1945), pp. 405–435.

[5] Herbert Robbins, abstract ("A one-sided confidence interval for an unknown distribution function,") *Ann. Math. Stat.*, Vol. 25 (1954), p. 409.

[6] N. V. Smirnov, "Approximate laws of distribution of random variables from empirical data," *Uspehi Matem. Nauk*, Vol. 10 (1944), pp. 179–206.

# NULL DISTRIBUTION OF THE HODGES BIVARIATE SIGN TEST

By Jerome Klotz[1]

*University of California, Berkeley*

**0. Summary.** This note presents the solution to the problem of obtaining the full null distribution for the bivariate sign test proposed by J. L. Hodges, Jr., in 1955 [1]. The partial solution of [1] is completed, and the table of [1] is extended to give the full distribution up to a sample size ($n$) of 30. In addition a partial table is included for sample size from 31 to 50.

**1. Introduction.** Using the notation given in [1], the problem is that of counting the number of cycles having a given value of $K$. This problem was solved in [1] only for the case $k < n/3$, or $n < 3h$ where $h = n - 2k$.

**2. Counting the cycles.** As stated in [1] the operation of rotation generates equivalence classes of cycles. We count the classes by selecting a representative member called a pattern. The number of cycles in each class is first determined. The total number of cycles for a given $k$ value is then obtained by summing these numbers over all patterns corresponding to $k$.

To every cycle corresponds a walk in the plane. A plus sign corresponds to a step in the $y$ direction, a minus sign to a step in the $x$ direction. Let us call a point $(x, y)$ a *departure point* for a path if it lies on the line $y = x$ (or $y = x + h$) and the path reaches the line $y = x + h$ ($y = x$) before returning to the line $y = x$ ($y = x + h$). Thus such points depend upon the given value of $h$ and the particular path. Further, let us call the path between consecutive departure points a *flight*.

**3. Specifying the patterns.** To every cycle corresponds the particular cycle called a pattern which is obtained from the first by rotation and has the following properties:

(i) The minimum number of minus signs above the diameter ($k$) is attained for this cycle.

(ii) The cycle starts with a plus and ends the $n$th step with a plus.

(iii) The first and hence $n$th points are departure points.

To see the existence of such a pattern for a given cycle, let the cycle be rotated until Condition (i) is satisfied. Condition (ii) must also be satisfied, otherwise, using the fact that diametrically opposed signs are opposite, rotation by one would decrease $k$ contrary to the initial rotation. If Condition (iii) is not satisfied, after an even number of steps the path from the first point (0, 0) returns to the line $y = x$. Thus, we can rotate the cycle so that the first point becomes a

Received February 16, 1959; revised April 17, 1959.

[1] This paper was prepared with the partial support of the Office of Ordnance Research U. S. Army under Contract DA-04-200-ORD-171, Task Order 3.

## TABLE OF $P[K \leqq k]$

$k$

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00000 | | | | | | | | | | | | | | | | |
| 2 | 1.00000 | | | | | | | | | | | | | | | | |
| 3 | .75000 | 1.00000 | | | | | | | | | | | | | | | |
| 4 | .50000 | 1.00000 | | | | | | | | | | | | | | | |
| 5 | .31250 | .93750 | 1.00000 | | | | | | | | | | | | | | |
| 6 | .18750 | .75000 | 1.00000 | | | | | | | | | | | | | | |
| 7 | .10938 | .54688 | .98438 | 1.00000 | | | | | | | | | | | | | |
| 8 | .06250 | .37500 | .87500 | 1.00000 | | | | | | | | | | | | | |
| 9 | .03516 | .24609 | .70312 | .99609 | 1.00000 | | | | | | | | | | | | |
| 10 | .01953 | .15625 | .52734 | .93750 | 1.00000 | | | | | | | | | | | | |
| 11 | .01074 | .09668 | .37598 | .80566 | .99902 | 1.00000 | | | | | | | | | | | |
| 12 | .00586 | .05859 | .25781 | .64453 | .96875 | 1.00000 | | | | | | | | | | | |
| 13 | .00317 | .03491 | .17139 | .48877 | .87280 | .99975 | 1.00000 | | | | | | | | | | |
| 14 | .00171 | .02051 | .11108 | .35547 | .73315 | .98437 | 1.00000 | | | | | | | | | | |
| 15 | .00092 | .01190 | .07050 | .24994 | .58319 | .91675 | .99994 | 1.00000 | | | | | | | | | |
| 16 | .00049 | .00684 | .04395 | .17090 | .44434 | .79980 | .99219 | 1.00000 | | | | | | | | | |
| 17 | .00026 | .00389 | .02698 | .11414 | .32684 | .66095 | .94551 | .99998 | 1.00000 | | | | | | | | |
| 18 | .00014 | .00220 | .01634 | .07471 | .23346 | .52295 | .84984 | .99610 | 1.00000 | | | | | | | | |
| 19 | .00007 | .00123 | .00978 | .04805 | .16264 | .39922 | .72450 | .96433 | .99999 | 1.00000 | | | | | | | |
| 20 | .00004 | .00069 | .00580 | .03044 | .11089 | .29572 | .59143 | .88737 | .99805 | 1.00000 | | | | | | | |
| 21 | .00002 | .00038 | .00340 | .01903 | .07420 | .21347 | .46575 | .77626 | .97666 | 1.00000 | 1.00000 | | | | | | |
| 22 | .00001 | .00021 | .00198 | .01175 | .04883 | .15068 | .35578 | .65057 | .91552 | .99901 | 1.00000 | | | | | | |
| 23 | .00001 | .00012 | .00115 | .00718 | .03167 | .10429 | .26474 | .52605 | .81834 | .98472 | 1.00000 | 1.00000 | | | | | |
| 24 | .00000 | .00006 | .00066 | .00434 | .02027 | .07094 | .19254 | .41259 | .70140 | .93665 | .99951 | 1.00000 | | | | | |
| 25 | | .00003 | .00038 | .00260 | .01282 | .04750 | .13723 | .31517 | .58020 | .85252 | .99000 | 1.00000 | 1.00000 | | | | |
| 26 | | .00002 | .00021 | .00155 | .00802 | .03137 | .09606 | .23525 | .46559 | .74495 | .95248 | .99975 | 1.00000 | | | | |
| 27 | | .00001 | .00012 | .00092 | .00497 | .02045 | .06616 | .17202 | .36390 | .62855 | .88028 | .99346 | 1.00000 | 1.00000 | | | |
| 28 | | .00001 | .00007 | .00054 | .00305 | .01318 | .04491 | .12350 | .27789 | 51460* | .78222 | .96436 | .99987 | 1.00000 | | | |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 29 | | .00000 | .00004 | .00031 | .00186 | .00841 | .03008 | .08722 | .20786 | .41040* | .67156 | .90281 | .99571 | 1.00000 | 1.00000 | | |
| 30 | | | .00002 | .00018 | .00112 | .00531 | .01991 | .06067 | .15263 | .31979* | .55963 | .81407 | .97327 | .99994 | 1.00000 | | |
| 31 | .00000 | .00000 | .00001 | .00010 | .00067 | .00332 | .01303 | .04163 | .11020 | | | | | | | | |
| 32 | | .00000 | .00001 | .00006 | .00040 | .00206 | .00750 | .02821 | .07837 | .18286 | | | | | | | |
| 33 | | | .00000 | .00003 | .00023 | .00127 | .00542 | .01890 | .05496 | .13469 | | | | | | | |
| 34 | | | .00000 | .00002 | .00014 | .00078 | .00344 | .01253 | .03805 | .09770 | .21371 | | | | | | |
| 35 | | | .00000 | .00001 | .00008 | .00047 | .00217 | .00822 | .02603 | .06987 | .16029 | | | | | | |
| 36 | | | .00000 | .00001 | .00005 | .00029 | .00136 | .00534 | .01761 | .04932 | .11836 | | | | | | |
| 37 | | | | .00000 | .00003 | .00017 | .00085 | .00345 | .01180 | .03440 | .08617 | .18663 | | | | | |
| 38 | | | | .00000 | .90002 | .00010 | .00052 | .00220 | .00783 | .02372 | .06191 | .14009 | | | | | |
| 39 | | | | .00000 | .00001 | .00006 | .00032 | .00140 | .00515 | .01619 | .04394 | .10366 | | | | | |
| 40 | | | | .00000 | .00001 | .00004 | .00020 | .00088 | .00336 | .01094 | .03084 | .07569 | .16260 | | | | |
| 41 | | | | | .00000 | .00002 | .00012 | .00055 | .00217 | .00733 | .02141 | .05460 | .12212 | | | | |
| 42 | | | | | .00000 | .00001 | .00007 | .00034 | .00140 | .00487 | .01472 | .03893 | .09052 | .18567 | | | |
| 43 | | | | | .00000 | .00001 | .00004 | .00021 | .00089 | .00321 | .01003 | .02746 | .06626 | .14138 | | | |
| 44 | | | | | | .00000 | .00003 | .00013 | .00056 | .00210 | .00677 | .01918 | .04795 | .10624 | | | |
| 45 | | | | | | .00000 | .00002 | .00008 | .00036 | .00136 | .00453 | .01327 | .03433 | .07885 | .16125 | | |
| 46 | | | | | | .00000 | .00001 | .00005 | .00022 | .00088 | .00301 | .00910 | .02433 | .05785 | .12272 | | |
| 47 | | | | | | .00000 | .00001 | .00003 | .00014 | .00056 | .00199 | .00619 | .01708 | .04198 | .09225 | .18158 | |
| 48 | | | | | | | .00000 | .00002 | .00009 | .00036 | .00130 | .00417 | .01188 | .03016 | .06854 | .13983 | |
| 49 | | | | | | | .00000 | .00001 | .00005 | .00023 | .00085 | .00279 | .00819 | .02146 | .05038 | .10635 | |
| 50 | | | | | | | .00000 | .00001 | .00003 | .00014 | .00055 | .00186 | .00561 | .01513 | .03665 | .07997 | .15743 |

* These values differ slightly from [1].

departure point without changing the number of plus or minus signs up. Conditions (i), (ii), and (iii) are now satisfied. We see that a pattern defined by the three conditions is not necessarily unique. Rotations of patterns taking departure points into like departure points may result in a different pattern. However, this need not concern us provided we count only one pattern for every class of patterns obtained by cyclic permutation. It may be noted that the definition of a pattern given here differs from that given in [1]. For example, [1] does not specify a pattern for cycles with alternating signs. However, under the restriction $k < n/3$, the two definitions are equivalent. For, under the restriction, a pattern satisfying the above conditions satisfies those of [1], and the pattern defined in [1] is unique.

**4. Counting the cycles for a pattern.** To count the number of cycles corresponding to a particular pattern we show that, to have less than $2n$ cycles for a pattern, the pattern must have an odd number (greater than one) of flights, which are all "equivalent." We specify two flights to be *equivalent* if they are identical, or if one can be obtained from the other by interchanging plus and minus signs (e.g., $++-++$ is equivalent ($\sim$) with $--+--$). Assume that after a rotation of less than $2n$ steps the pattern repeats itself—we take a pattern as a starting point for convenience. We must have departure points going into departure points of the same kind and hence the rotation consists of an even number of flights. Let us denote the flights by $\alpha_i$ and suppose there are $2t + 1$ flights in the pattern $i = 1, 2, \cdots, 2t + 1$. Next represent the cycle by $(\alpha_1, \alpha_2; \alpha_3, \alpha_4; \cdots; \alpha_{2t+1} \mid \alpha_{2t+2}; \cdots, \alpha_{4t+2})$. After a rotation which repeats the pattern—say a rotation of $2p$ flights we obtain

$$\begin{array}{l} \alpha_1 \sim \alpha_{2+p1} \sim \alpha_{4p+1} \sim \cdots \\ \vdots \\ \alpha_{2p} \sim \alpha_{4p} \sim \cdots \end{array}$$

where $\alpha_m = \alpha_{m(\mathrm{mod}\, 4t+3)}$. From the symmetry (diagonally opposed signs are opposite) we have $\alpha_1 \sim \alpha_{2t+2}, \alpha_2 \sim \alpha_{2t+3}, \cdots, \alpha_{2t+1} \sim \alpha_{4t+2}$. Solving the system, we obtain $\alpha_1 \sim \alpha_2 \sim \cdots \alpha_{2t+1} \cdots$. Thus for this case we have $2n/2t + 1$ cycles for the pattern and otherwise $2n$ cycles.

**5. Counting the patterns.** To count the patterns we count them according to their number of flights—one, three, five, etc. For $(2l + 1)h \leqq n < (2l + 3)h$ we will have patterns with $1, 3, 5, \cdots, 2l + 1$ flights. For the case of only one flight, [1] gives the formula for $2^n P[K = k] = 2nm_h(n)$. $m_h(n)$ is the number of ways of going from $(0, 0)$ to $(k, n - k)$ hitting the line $y = x + h$ only at the $n$th step—the gambler's ruin problem (see [1]). Generalizing, we obtain the formula for the case $(2l + 1)h \leqq n < (2l + 3)h$, where we have up to $2l + 1$ flights

$$2^n P[K = k] = 2nm_h(x) + \frac{2n}{3}\left[\binom{3}{1,\,1} \sum_{\substack{n_1 < n_2 < n_3 \\ n_1+n_2+n_3=n}} \prod_{i=1}^{3} m_h(n_i) + \binom{3}{2} \sum_{\substack{n_1 = n_2 \neq n_3 \\ n_1+n_2+n_3=n}} \prod^{3} m_h(n_i)\right]$$

$$+ I_{[3a=n]} \frac{2n}{3}\left[\binom{3}{1,1}\binom{m_h(a)}{3} + 2\binom{3}{2}\binom{m_h(a)}{2} + \binom{m_h(a)}{1}\right] + \cdots$$

$$+ \frac{2n}{2l+1}\left[\binom{2l+1}{1,1,\cdots,1}\sum_{n_1<\cdots<n_{2l+1}}\prod_{i=1}^{2l+1} m_h(n_i) + \binom{2l+1}{2,1,\cdots,1}\sum\prod m_h(n_i)\right.$$

$$+ \cdots + \binom{2l+1}{r_1, r_2, \cdots, r_{t-1}}\sum\prod m_h(n_i) + \cdots$$

$$\left. + \binom{2l+1}{2l}\sum_{\substack{n_1=\cdots=n_{2l+1}\\ \sum n_i=n}}\prod m_h(n_i)\right] + I_{[(2l+1)q=n]}\frac{2n}{2l+1}\left[\binom{2l+1}{1,1,\cdots,1}\binom{m_h(q)}{2l+1}\right.$$

$$+ 2l\binom{2l+1}{2,1,\cdots,1}\binom{m_h(q)}{2l} + \cdots + (t)_{p-1}\binom{2l+1}{r_1, r_2, \cdots, r_{t-1}}\binom{m_h(q)}{t} + \cdots$$

$$\left. \cdots\cdots + \binom{m_h(q)}{1}\right]$$

where $p$ is the number of different $r_i$; $\sum_{i=1}^{t} r_i = n$; $a, \cdots, q$; $r_i$ are integers; $\binom{n}{k_1 \cdots k_{t-1}} = n!/k_1! \cdots k_t!$ is the multinomial coefficient; and $I$ is an indicator function.

The preceding table completes the table of [1] and gives values of $P[K \leq k]$ to $5D$ for all values of $k$ and $n = 1(1)30$. Further the table gives values of $P$ to $5D$ for $k$ up to a value which makes $P$ just greater than 10 per cent and $n = 31(1)50$.

**6. Acknowledgment.** I wish to thank Professor J. L. Hodges, Jr. for suggesting a method of attack that led to the solution.

## REFERENCE

[1] J. L. Hodges, Jr., "A bivariate sign test," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 523–527.

# EXACT NONPARAMETRIC TESTS FOR RANDOMIZED BLOCKS

By John E. Walsh[1]

*Lockheed Aircraft Corporation*

**1. Summary.** A class of nonparametric procedures for testing the statistical identity of treatments in randomized block experiments is suggested and discussed. The suggested procedures are squarely based on experimental within-block randomizations, and they may be chosen so as to have special power against particular alternatives. The blocks are assumed to be statistically independent but no assumption is made concerning dependence within the various blocks. The basic idea is to obtain from each block a statistic that is, under the null hypothesis, symmetrically distributed about zero and then to apply to the set of these statistics a nonparametric test of symmetry about zero. The observational data can be of any quantitative type.

**2. Introduction.** This paper considers experimental designs that are laid out in statistically independent blocks. If care is exercised, the blocks can usually be separated enough in distance, time, etc., to warrant the assumption of statistical independence.

Within a block, the assignment of the treatments investigated in that block can be of either a balanced or an unbalanced nature. For a given design, some blocks might be balanced and others unbalanced. The within-block assignments of treatments to locations are determined by a set of independent randomization processes as follows: the treatments of each block are partitioned into disjoint classes, to each class there is assigned a set of eligible locations within the block, and the assignments of treatments within a class to their eligible locations, for some classes, those of type A, are strictly random (all assignments equally likely), possibly dependent from class to class but independent from block to block. A block always contains at least one class of type A and each of these contains at least two treatments. For the remaining classes, those of type B, assignment to location may be random or fixed. The partitioning scheme, which may vary from block to block, is selected on the basis of the null hypothesis and the alternative hypotheses being investigated.

The most elementary type of situation considered is that where, for each block, the treatments (at least two per block) are not partitioned. Then all the classes (one per block) are of type A and the null hypothesis asserts that, for each block, the joint distribution of the observations is invariant under all permutations of the names of treatments within each block. Also, within a block, the locations are eligible for all the treatments and are randomly assigned to these treatments.

This elementary situation can be generalized in several respects through the

Received September 30, 1957; revised March 28, 1958.

[1] Now with the System Development Corporation, Santa Monica, California.

use of partitioning. First, various combinations of treatments can be selected for comparison within a block. Second, the eligible locations associated with the classes of the partition can be chosen in many ways. Finally, the part of the null hypothesis that pertains to a given block need not consider all the treatments of this block. In fact, only the treatments of the partition classes of type A are considered. The null hypothesis asserts that the joint distribution of the observations for a block is invariant under all permutations of the names of treatments within a partition class for all the classes of type A. That is, excluding partition classes of type B, under the null hypothesis the treatments within a partition class have identical joint probability properties.

The procedure of including treatments in a block which are not considered in the part of the null hypothesis pertaining to this block serves a useful purpose. For the situations of this paper, a treatment is included in the experiment for one or both of two reasons. First, the question of whether this treatment is identical with a specified one or more other treatments can be of interest. This type of relation is considered in selection of the null hypothesis $H_0$. Second, there can be interest in a given form of interrelation that might exist between this treatment and specified other treatments when the null hypothesis is false. This second type of relation is considered in identifying the alternative hypotheses that are of principal interest. For each block, those treatments which are included exclusively for the second reason are placed in one or more partition classes of type B and are not considered in the part of the null hypothesis that is associated with this block. The reason for using more than one partition class for these treatments (e.g., a separate partition class for each treatment) is that they may not have the same set of eligible locations.

The choice of the eligible locations for the various classes of the partition is at the discretion of the experimenter. Often this freedom of choice in specifying eligible locations can be exploited to obtain a more efficient experiment. As an example, for the treatments of some partition classes, location in one part of the block may be more important than location in other parts, because of a special condition that exists in this part. There is great freedom in selecting eligible locations for treatments, subject to the condition that the size of each set of eligible locations is at least as great as the number of treatments that could be assigned to this set. In particular, two or more classes of the partition might have over-lapping sets of eligible locations. For this case, it is convenient, but not necessary, to require that if any two classes of the partition are to have at least one eligible location in common they must have the same set of eligible locations. Then, when two or more of the partition classes have the same set of eligible locations, their treatments can be handled as a group in performing the random assignment to eligible locations. This "grouping and then assigning" procedure greatly simplifies the random assignment scheme for situations of this nature. If desired, a specified location can be assigned to each treatment of a partition class of type B.

To perform the test, a statistic is specified for each block. This statistic de-

pends on all the treatments for this block but not on those for any of the other blocks. These statistics are chosen so that they have symmetrical distributions about zero when the null hypothesis is true. They are also chosen so that the test is sensitive to the alternative hypotheses that are emphasized. The forms of the statistics can vary from block to block and this freedom can sometimes be exploited by tailoring the statistics to the special situations that exist for the blocks. Because of the broad nature of the situations considered, no generally applicable rules can be stated for choosing the block statistics so as to emphasize the alternative hypotheses of major interest. However, in many cases a reasonable selection can be made on an intuitive basis. Some examples of the selection of block statistics are given in Section 4. Since the observations are independent between blocks, the block statistics are independent and also, under the null hypothesis, have symmetrical distributions about zero. Consequently the null hypothesis can be tested by use of an appropriate nonparametric test of symmetry about zero. References to some nonparametric tests of symmetry about zero are given in Section 3.

No quantitative attempt is made to evaluate the efficiencies of the tests that can be obtained on the basis of this paper. The great freedom allowed in selecting the treatment partition classes, the eligible locations, and the block statistics, combined with the myriad of possible alternative hypotheses and the different kinds of tests that could be used, make such an investigation infeasible. Qualitative considerations, however, hint that in many cases the efficiency should be reasonably high if the eligible assignment locations and the block statistics are chosen so that the alternative hypotheses of major interest are emphasized. For example, if the normality model for experimental design holds and the treatment comparisons are linear, the best test is that based on the appropriate $t$-statistic. A situation of this nature was examined in [1], under conditions that represent a special case of the results of this paper. The tests based on the block statistics used in [1] were found to have efficiencies that are only in the neighborhood of 60–70 percent for the case of normality and linear comparisons. However, if the most appropriate treatment comparison for the alternative hypotheses of interest is not linear, suitable selection of the block statistics so as to emphasize these alternative hypotheses may furnish the basis for nonparametric tests that are much more efficient than the best $t$-tests based on linear comparisons. Of course, if the block statistics and the eligible locations are poorly chosen, a test of this type can have a very low efficiency.

It is no loss of generality to suppose, for purposes of formal theory, that the treatments of a block are possibly different, or at least have different names; also to suppose that different blocks can contain different treatments. Situations where treatments are replicated or where the same treatments occur in several blocks represent special cases of this general situation.

The null hypothesis of treatment equivalence for specified treatment partitions can be generalized. Instead of specifying that, for the partition classes whose treatments are named in $H_0$ (i.e., class A), the treatments of a partition

class have identical probability properties, the null hypothesis could assert that this is the case if specified transformations are made of these values. By use of transformations, the classes of null hypotheses that are available for consideration and of alternative hypotheses that are emphasized by elementary form block statistics can be greatly extended.

In Section 3 the permissible forms for the block statistics, verification of their properties under the null hypothesis, and a statement of how the test makes use of these statistics are presented. Section 4 is titled Block Statistic Selection. Two examples are given to illustrate the intuitive selection of block statistics so as to emphasize specified alternative hypotheses.

**3. Results.** The principal purpose of this section is to show that a properly chosen block statistic is symmetrically distributed about zero under $H_0$. Consequently, all the notation occurring in the derivation applies to an arbitrary but specified block. The random method used to assign the treatments of a partition to their eligible locations is described in Section 2 and is not repeated here.

Suppose that treatments $1, 2, \cdots, I$ occur in the block, and that these are partitioned into $T + 1$ classes. The first $T$ of these classes are of type A and the last of type B. (If there are several partition classes of type B, nothing is lost by throwing them together and working conditionally on whatever random assignments to location may have been made for such type B classes.) The $t$th set contains $k(t) - k(t - 1)$ treatments, with $k(0) = 0$ and $k(T + 1) = I$. The treatments for the $t$th set are denoted by

$$i_{k(t-1)+1}, i_{k(t-1)+2}, \cdots, i_{k(t)} \qquad (t = 1, \cdots, T + 1).$$

The partitioning is done so that sets $1, \cdots, T$ are the partition classes which are used in the part of the null hypothesis pertaining to this block (i.e., class A), while the remaining set contains all the treatments that are in partition classes which do not appear in $H_0$ (class B). In terms of this notation, the null hypothesis associated with this block asserts that all the treatments of the $t$th set have identical joint probability properties with respect to the experiment for $t = 1, \cdots T$.

Let the random variable $y(i)$ represent the observable result for the $i$th treatment ($i = 1, \cdots, I$), where the probability effects from the randomization and the experimentation are combined to obtain the joint distribution of $y(1), \cdots, y(I)$. For $1 \leqq t \leqq T$, let $\phi_t$ denote an arbitrary permutation of the numbers $i_{(k-1)+1}, \cdots, i_{k(t)}$; also let $\phi_{T+1}$ be the identity transformation for $i_{k(T)+1}, \cdots, i_I$. Use

$$F\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T + 1\}$$

to denote the joint cumulative distribution function (cdf) for $y(1), \cdots, y(I)$. Then, on the basis of the randomization scheme and the null hypothesis,

$$\begin{aligned} &F\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T + 1\} \\ &\qquad \equiv F\{y[\phi_t(i_{k(t-1)+1})], \cdots, y[\phi_t(i_{k(t)})]; 1 \leqq t \leqq T + 1\}. \end{aligned} \tag{1}$$

That is, under $H_0$, the joint cdf of $y(1), \cdots, y(I)$ is invariant under all possible permutations within each of the $T$ sets of treatments that are considered in the null hypothesis. No moments of any order are assumed to exist for the $y(i)$.

A block statistic is a function of $y(1), \cdots, y(I)$ which is denoted by

$$g\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T + 1\}.$$

This function, which is chosen so as to not be identically zero for all values of the $y(i)$, is required to have the property that there exists a set of permutations $\phi_1, \cdots, \phi_T, \phi_{T+1}$, where $\phi_{T+1}$ is the identity permutation, such that

$$g\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T + 1\}$$
$$\equiv -g\{y[\phi_t(i_{k(t-1)+1})], \cdots, y[\phi_t(i_{k(t)}]; 1 \leqq t \leqq T + 1\}.$$

But, on the basis of relation (1),

$$g\{y[i_{k(t-1)+1}], \cdots, y[i_{k(t)}]; 1 \leqq t \leqq T + 1\}$$

and

$$g\{y[\phi_t(i_{k(t-1)+1})], \cdots, y[\phi_t(i_{k(t)})]; 1 \leqq t \leqq T + 1\}$$

have the same distribution. Thus $-g$ has the same distribution as $g$ if the null hypothesis holds; consequently, under $H_0$, the block statistic $g$ has a probability distribution that is symmetrical about zero.

Since, by hypothesis, the observations are statistically independent between blocks, the block statistics are a set of independent random variables with distributions that are symmetrical about zero if the null hypothesis is true. A wide variety of nonparametric procedures are available for testing the symmetry of populations about zero. These include the signed-rank test of Wilcoxon [2], [3], [4], [5], the Fisher test [6], Nair's test [7], a comprehensive set of tests by Hemelrijk [8] and by van Eeden and Benard [9], and the results of [10], [11]. If the distributions of the block statistics are not all continuous, tests based on the assumption of continuity can be validly used by appropriate randomization of ties. Alternately, some of the tests are valid for both discrete and continuous populations (see, e.g., [6], [7], [8], [9]).

The efficiency of this testing procedure depends on the test used, the forms of the block statistics, the partitioning scheme, and the choice of eligible locations for treatments. In particular, the forms of the block statistics have a strong influence on which alternative hypotheses are emphasized. The next section considers intuitively the problem of choosing the forms of the block statistics so as to emphasize specified types of alternative hypotheses.

**4. Block statistic selection.** The great freedom in selecting the forms for the block statistics allows so many types of situations to arise that no general rule for the selection of these statistics seems to be available. The alternative hypotheses which are eligible for consideration are of such a wide class that determination of a general method of selecting a block statistic so as to emphasize an arbitrary

but specified alternative hypothesis (hypotheses) does not appear to be feasible. However, a reasonable (but not necessarily preferable) selection can often be made on the basis of judgment combined with intuitive considerations. Two examples are given which illustrate the intuitive method of selecting block statistics and which are somewhat typical of situations of practical interest. In these examples, the same form is considered to be usable for all blocks. However, since the considerations are on the basis of a single block, these considerations also apply to cases where the forms may change from block to block.

*First example*: Let $I = 8$ and suppose that the null hypothesis asserts that treatment 1 is equivalent to treatment 2 and that treatments 3–6 are equivalent. Treatments 7 and 8 do not occur in the statement of the null hypothesis. The three alternative hypotheses of principal interest are

$H_1$ : The value of treatment 1 tends to be larger than that of treatment 2, but small deviations are not important.

$H_2$ : The average of the values of treatments 3 and 4 tends to be smaller than the average of the values of treatments 5 and 6.

$H_3$ : The value for treatment 1 minus that for treatment 2 tends to be negative and simultaneously the average of the values of treatments 3–8 tends to exceed 10.

If all of $H_1$–$H_3$ hold, or if at least one holds in a strong fashion and neither of the one-sided $H$'s holds strongly in a negative sense, it is highly desirable that the null hypothesis be rejected.

For this case, use of the function

$$\{y[1] - y[2]\}^3 - \tfrac{1}{2}\{y[3] + y[4] - y[5] - y[6]\}$$
$$- \tfrac{1}{6}\{y[3] + \cdots + y[8] - 60\}\ sgn\,\{y[1] - y[2]\}$$

for $g$, combined with an appropriate one-sided test for symmetry about zero (which is sensitive to large positive values of the variable) might be satisfactory. The first term accounts for the alternative $H_1$, the second term for $H_2$, and the third term for $H_3$. The permutations

$$\phi_1 : 1 \leftrightarrow 2 \qquad \phi_2 : 3 \leftrightarrow 5 \quad \text{and} \quad 4 \leftrightarrow 6$$

result in a change of sign for $g$.

*Second example*. Let $I = 4$. The null hypothesis asserts that all four treatments are equivalent. The alternative hypothesis of principal interest is

$H_1$ : The sign of the value of treatment 1 minus that of treatment 2 tends to be the same as that of the value of treatment 3 minus that of treatment 4. Also the magnitude of the difference involving treatments 1 and 2 tends to exceed that of the difference involving treatments 3 and 4.

For this case, use of the function

$$\{y[1] - y[2]\}/\{y[3] - y[4]\}$$

for $g$, combined with an appropriate one-sided test of symmetry about zero (which is sensitive to large positive values of the variables), would appear to

be suitable. The permutation

$$\phi_1 : 1 \leftrightarrow 2, 3 \leftrightarrow 3, \quad \text{and} \quad 4 \leftrightarrow 4$$

results in a change of sign for $g$.

## REFERENCES


[1] JOHN E. WALSH, "Some nonparametric tests for Student's hypothesis in experimental designs," *Jour. Amer. Stat. Assoc.*, Vol. 47 (1952), pp. 401–415.

[2] FRANK WILCOXON, "Individual comparisons by ranking methods," *Biom. Bull.*, Vol. 1 (1945), pp. 80–83.

[3] FRANK WILCOXON, "Individual comparisons of grouped data by ranking methods," *Jour. Econ. Entomology*, Vol. 39 (1946), pp. 269–270.

[4] FRANK WILCOXON, "Probability tables for individual comparisons by ranking methods," *Biometrics*, Vol. 3 (1947), pp. 119–122.

[5] FRANK WILCOXON, *Some Rapid Approximate Statistical Procedures*, American Cyanamid Co., Stamford Research Laboratories, 1949.

[6] R. A. FISHER, *The Design of Experiments* (5th Ed.), Hafner Publishing Co., Inc., N. Y., 1949, pp. 43–47.

[7] K. R. NAIR, "The median in tests by randomization," *Sankhyā*, Vol. 4 (1940), pp. 543–550.

[8] J. HEMELRIJK, "A family of parameterfree tests for symmetry with respect to a given point I, II," *Proc. Kon. Ned. Akad. v. Wet., Ser. A*, Vol. 53 (1950), pp. 945-955 and pp. 1186–1198; also *Indag. Math.*, Vol. 12 (1950), pp. 340–350 and 419–431.

[9] CONSTANCE VAN EEDEN AND A. BENARD, "A general class of distribution-free tests. for symmetry containing the tests of Wilcoxon and Fisher I, II, III," *Proc. Kon. Ned. Akad. v. Wet., Ser. A*, Vol. 60 (1957), pp. 381–408; also *Indag. Math.*, Vol. 19 (1957).

[10] JOHN E. WALSH, "Some significance tests for the median which are valid under very general conditions," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 64–81.

[11] JOHN E. WALSH, "Applications of some significance tests for the median which are valid under very general conditions," *Jour. Amer. Stat. Assoc.*, Vol. 44 (1949), pp. 342–355.

# A GENERALISATION OF PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS[1]

By B. V. Shah

*University of Bombay*

**1. Introduction.** The partially balanced incomplete block (PBIB) designs were first defined by Bose and Nair [2] in 1939. Later on, in 1942, Nair and Rao [7] generalised the original definition to include some confounded factorial designs as well as many others in the class of PBIB designs. The class of PBIB designs was found to include most of the designs used in practice. In 1946, Harshberger [4] presented triple rectangular lattices, and Nair [6] proved that these designs were not in general PBIB designs, but that the duals of these designs were PBIB designs. So it was found that, except for the intra-inter-group balanced designs given by Nair and Rao [8], almost all the designs so far proposed, with limited number of distinct variances for elementary treatment comparisons, were either PBIB designs or duals of PBIB designs. Yet a need was felt to find a more general class of designs. In an attempt to find out why the PBIB designs with $m$ associate classes have $m$ distinct types of treatment comparisons, I came across a more general class of designs, which is given in this paper. The arguments which led to this generalisation are also put forward.

**2. Notation.** Let there be $v$ treatments, each replicated $r$ times in $b$ blocks of $k$ plots each. Let $\mathbf{N} = [n_{ij}]$ $(i = 1, 2, \cdots, v; j = 1, 2, \cdots, b)$ be the incidence matrix of the design, where $n_{ij}$ is equal to the number of times the $i$th treatment occurs in the $j$th block. It is assumed that $n_{ij}$ is 0 or 1. The assumed model is

$$y_{ij} = \mu + \beta_j + \hat{t}_i + \epsilon_{ij}, \tag{2.1}$$

where $y_{ij}$ is the yield of the plot in the $j$th block to which the $i$th treatment is applied, $\mu$ is the general effect, $\beta_j$ is the effect of the $j$th block, $\hat{t}_i$ is the effect of the $i$th treatment and $\epsilon_{ij}$'s are independent normal variates with mean 0 and variance $\sigma^2$. Let $T_i$ be the total yield of all the plots having the $i$th treatment, $B_j$ be the total yield of all the plots of the $j$th block and $t_i$ be a solution for $\hat{t}_i$ in the normal equations. Further denote the column vectors $\{T_1, T_2, \cdots, T_v\}$, $\{B_1, B_2, \cdots, B_b\}$, $\{t_1, t_2, \cdots, t_v\}$ and $\{\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_v\}$ by $\mathbf{T}$, $\mathbf{B}$, $\mathbf{t}$ and $\hat{\mathbf{t}}$ respectively. It is well known that the reduced normal equations for the intra-block estimates of the treatment contrasts are

$$\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}, \tag{2.2}$$

where

$$\mathbf{Q} = \mathbf{T} - \frac{1}{k}\mathbf{NB} \tag{2.3}$$

Received January 6, 1959; revised June 16, 1959.

[1] This work was supported by a Government of India Research Training Scholarship.

and

$$\mathbf{C} = r\mathbf{I}(v) - \frac{1}{k}\,\mathbf{NN}', \tag{2.4}$$

where $\mathbf{I}(v)$ is the $v \times v$ Identity matrix. The matric $\mathbf{C}$ defined in (2.4) will be called the $\mathbf{C}$-matrix of the design. Denote by $\mathbf{E}(m, n)$ the $m \times n$ matrix with all its elements equal to 1.

LEMMA 2.1: *If the design is connected, the matrix* $\mathbf{C} + a\mathbf{E}(v, v)$ *is non-singular, where a is any non-zero real number and* $\hat{\mathbf{t}} = [\mathbf{C} + a\mathbf{E}(v, v)]^{-1}\mathbf{Q}$ *is a solution of the equation* $\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}$.

PROOF: Let $\theta_1, \theta_2, \cdots, \theta_v$ be the canonical roots of the $\mathbf{C}$-matrix and let $\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_v$ be the corresponding canonical vectors. It is well known that the $\mathbf{C}$-matrix has one root 0 and that the corresponding canonical vector is $(v)^{-\frac{1}{2}}\mathbf{E}(v, 1)$; denote these by $\theta_1$ and $\mathbf{l}_1$ respectively. Then

$$\mathbf{C} = \sum_{i=2}^{v} \theta_i \mathbf{l}_i \mathbf{l}_i' \tag{2.7}$$

and

$$\mathbf{C} + a\mathbf{E}(v, v) = \sum_{i=2}^{v} \theta_i \mathbf{l}_i \mathbf{l}_i' + av\, \mathbf{l}_1 \mathbf{l}_1'. \tag{2.8}$$

Since the design is connected, the rank of the $\mathbf{C}$-matrix is $v - 1$, and therefore none of the $\theta_i$'s except $\theta_1$ is 0. Hence from (2.8) it follows that the matrix $\mathbf{C} + a\mathbf{E}(v, v)$ is non-singular and

$$[\mathbf{C} + a\mathbf{E}(v, v)]^{-1} = \sum_{i=2}^{v} \frac{1}{\theta_i} \mathbf{l}_i \mathbf{l}_i' + \frac{1}{av}\, \mathbf{l}_1 \mathbf{l}_1' \tag{2.9}$$

Also

$$\mathbf{C}[\mathbf{C} + a\mathbf{E}(v, v)]^{-1} = \sum_{i=2}^{v} \mathbf{l}_i \mathbf{l}_i' = \mathbf{I}(v) - \frac{1}{v}\,\mathbf{E}(v, v). \tag{2.10}$$

Hence, since $\sum Q_i = 0$, $[\mathbf{C} + a\mathbf{E}(v, v)]^{-1}\mathbf{Q}$ is a solution for $\hat{\mathbf{t}}$ in the equation $\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}$.

LEMMA 2.2: *If* $\hat{\mathbf{t}} = \mathbf{AQ}$ *is a solution of* $\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}$, *then* $[\mathbf{A} + a\mathbf{E}(v, v)]\mathbf{Q}$ *is also a solution of* $\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}$.

**3. PBIB designs.** An incomplete block design is said to be partially balanced (PBIB) if it satisfies the following conditions (Bose and Shimamoto [3]):

(i) The experimental material is divided into $b$ blocks of $k$ plots each, different treatments being applied to the plots in the same block.

(ii) There are $v$ treatments each of which occurs in $r$ blocks.

(iii) There can be established relations of association between any two treatments satisfying the following requirements:

(a) Two treatments are either 1st, 2nd, $\cdots$, or $m$th associates.

(b) Each treatment has exactly $n_i$ $i$th associates ($i = 1, 2, \cdots, m$).

(c) Given any two treatments which are $i$th associates, the number of treatments common to the $j$th associates of the first and the $k$th associates of the second is $p^i_{jk}$ and is independent of the pair of treatments with which we start. Also $p^i_{jk} = p^i_{kj}$.

(iv) Two treatments which are the $i$th associates occur together in exactly $\lambda_i$ blocks.

Now further define each treatment to be its own 0th associate and the 0th associate of no other treatment. We may thus consistently write

$$(3.1) \qquad \lambda_0 = r, \qquad n_0 = 1, \qquad p^0_{st} = \delta_{st} n_s, \qquad p^t_{0s} = p^t_{s0} = \delta_{st},$$

where $\delta_{ij}$ is the Kronecker delta which is defined for all pairs of natural numbers $i, j$, as $\delta_{ij} = 1$, if $i = j$; and $\delta_{ij} = 0$, if $i \neq j$. Then the relations between the parameters are

$$bk = vr, \qquad \sum_{i=0}^{m} n_i = v,$$

$$(3.2) \qquad \sum_{i=0}^{m} n_i \lambda_i = rk, \qquad \sum_{k=0}^{m} p^j_{ik} = n_i,$$

$$n_i p^i_{jk} = n_j p^j_{ik} = n_k p^k_{ij}, \qquad i, j, k = 0, 1, \cdots, m.$$

Now consider $v(v + 1)/2$ treatment pairs $(i, j)$ $(i, j = 1, 2, \cdots, v)$, assuming that $(i, j)$ is identical with $(j, i)$. Partition them into $(m + 1)$ disjoint classes and corresponding to the $t$th class ($t = 0, 1, \cdots, m$), define the $v \times v$ matrix $\mathbf{B}_t = [B^t_{ij}]$, where $B^t_{ij} = 1$, if the pair $(i, j)$ belongs to the $t$th class and $B^t_{ij} = 0$ otherwise. The classes can be called the association classes and the corresponding matrices, the association matrices. As there is one to one correspondence between the association classes and matrices defined above, either of them will uniquely determine the other. It can be seen that each $\mathbf{B}_t$ is symmetric. Since every pair must belong to one of the association classes, it is obvious that

$$(3.3) \qquad \sum_{i=0}^{m} \mathbf{B}_i = \mathbf{E}(v, v).$$

THEOREM 3.1: *The necessary and sufficient conditions, that* $(m + 1)$ *association matrices* $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$ *determine an association scheme for an* $m$ *associate class* PBIB *design, are that*

$$(3.4) \qquad \mathbf{B}_0 = \mathbf{I}(v),$$

*and*

$$(3.5) \qquad \mathbf{B}_t \mathbf{B}_x = \sum_{i=0}^{m} p^i_{tx} \mathbf{B}_i, \qquad t, x = 0, 1, \cdots, m.$$

The proof of the above theorem follows immediately from the definition of a PBIB design given by Bose and Shimamoto [3].

The idea of association matrices was also developed by Bose and Mesner [1] independently and became available after submission of the manuscript of this paper. The reader may note that the concept introduced by Bose and Mesner is slightly different from one given here. The idea of association classes and matrices as introduced by the former is confined only to PBIB designs, whereas, my interest being the generalisation of PBIB designs, the association classes and matrices are defined in terms of partitioning of $v(v + 1)/2$ combinations of $v$ objects (an object may occur more than once) taken two at a time, into $(m + 1)$ mutually exclusive and exhaustive classes. Theorem 3.1 gives a set of necessary and sufficient conditions for such a scheme of partitioning to be an association scheme of a PBIB. Lemma 3.1 of Bose and Mesner [1] proves the necessary part of the condition; the sufficiency is proved by Lemma 5.1 of [1].

Before deriving further results, it is necessary to prove the following matrix theorem.

THEOREM 3.2: *If* $\mathbf{A}$ *is a* $v \times v$ *positive definite matrix, such that all the non-negative integral powers of* $\mathbf{A}$ *are of the form*

$$\mathbf{A}^N = \sum_{i=0}^{m} u_{Ni}\mathbf{B}_i, \qquad N = 0, 1, 2, \cdots, \tag{3.6}$$

*where* $u_{Ni}$ *are scalar constants and* $\mathbf{B}_i$ *are fixed* $v \times v$ *matrices and* $\mathbf{A}^0$ *means* $\mathbf{I}(v)$, *then the matrix* $\mathbf{A}^{-1}$ *must also be of the form* $\sum d_i\mathbf{B}_i$, *where* $d_i$ *are scalar constants.*

PROOF: Let $\theta$ be the maximum of the canonical roots of the matrix $\mathbf{A}$. Then the canonical roots of the matrix $\mathbf{B} = \mathbf{I}(v) - \{1/(\theta + 1)\}\mathbf{A}$ lie within the range 0 and 1. Now consider the series

$$\mathbf{D} = \sum_{N=0}^{\infty} \mathbf{B}^N. \tag{3.7}$$

The above series converges because the series $\sum x^N$ converges for $-1 < x < 1$ and the canonical roots of $\mathbf{B}$ lie within the range (Macdufee [5]). Also it can be shown that

$$\mathbf{AD} = (\theta + 1)\mathbf{I}(v) = \mathbf{DA}, \tag{3.8}$$

hence

$$\mathbf{A}^{-1} = \frac{1}{\theta + 1}\mathbf{D}. \tag{3.9}$$

Now since every power of $\mathbf{A}$ is a linear combination of matrices $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$, the same is true for every power of $\mathbf{B}$ and hence $\mathbf{D}$ is also a linear combination of the matrices $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$.

COROLLARY 3.2.1: *If there exist matrices* $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$, *such that* $\mathbf{I}(v)$, $\mathbf{E}(v, v)$, *and all the positive integral powers of the* $\mathbf{C}$*-matrix of a connected design are linear combinations of the matrices* $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$, *then there exists a solution* $\hat{\mathbf{t}} = \mathbf{AQ}$ *of the equation* $\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}$, *such that the matrix* $\mathbf{A}$ *is a linear combination of the matrices* $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$, *and also*

$$\mathbf{AC} = \mathbf{CA} = \mathbf{I}(v) - \frac{1}{v}\,\mathbf{E}(v, v). \tag{3.10}$$

The proof of Corollary 3.2.1 follows immediately from Theorem 3.2 and Lemma 2.1.

The **C**-matrix of a PBIB design can be written in the form

$$\mathbf{C} = \frac{r(k-1)}{k}\,\mathbf{B}_0 - \sum_{i=1}^{m} \frac{\lambda_i}{k}\,\mathbf{B}_i, \tag{3.11}$$

where $\mathbf{B}_t$ is the association matrix corresponding to the $t$th associate class ($t = 0, 1, \cdots, m$). Using relation (3.5) and mathematical induction, it can be proved that all the powers of **C** are linear combinations of $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$, also $\mathbf{I}(v) = \mathbf{B}_0$ and $\mathbf{E}(v, v) = \sum_0^m \mathbf{B}_i$. Hence, by Corollary 3.2.1, it follows that a solution $\hat{\mathbf{t}} = \mathbf{AQ}$ exists, such that

$$\mathbf{A} = \sum_{i=0}^{m} d_i\mathbf{B}_i. \tag{3.12}$$

With a little algebra, it can be shown that the $d_i$'s are the solutions of the equations

$$\begin{aligned} \sum_{i=0}^{m}\sum_{j=0}^{m} p_{ij}^{l} c_i\, d_j &= 1 - \frac{1}{v}, && \text{if } l = 0; \\ &= -\frac{1}{v}, && \text{if } l = 1, 2, \cdots, m, \end{aligned} \tag{3.13}$$

where

$$c_0 = \frac{r(k-1)}{k}, \qquad c_i = -\frac{\lambda_i}{k}, \qquad i = 1, 2, \cdots, m. \tag{3.14}$$

Since the $m + 1$ equations in (3.13) are not independent, any $m$ of them can be taken and solved with an additional convenient restriction like $\sum d_i = 0$, or, for some $j$, $d_j = 0$. It can be verified that the solutions obtained by taking $d_j = 0$ will be identical with those obtained by Bose and Nair [2].

## 4. Restrictions on association matrices.

LEMMA 4.1: *If* $\mathbf{C} = \sum_{i=0}^{m} c_i\mathbf{B}_i$ *and if*

$$\mathbf{B}_t\mathbf{B}_x + \mathbf{B}_x\mathbf{B}_t = 2\sum_{i=0}^{m} q_{tx}^{i}\mathbf{B}_i, \tag{4.1}$$

*for all* $x, t = 0, 1, \cdots, m$, *then*

$$\mathbf{C}^N = \sum_{i=0}^{m} u_{Ni}\mathbf{B}_i, \tag{4.2}$$

*for all positive integral values of* $N$.

PROOF: The theorem is true for $N = 1$. Assuming the result to be true for $N$, it can be proved for $N + 1$ as follows:

Since a matrix commutes with its powers,

$$\mathbf{C}^{N+1} = \mathbf{C}^N\mathbf{C} = \mathbf{C}\mathbf{C}^N. \tag{4.3}$$

Therefore

$$\mathbf{C}^{N+1} = \tfrac{1}{2}(\mathbf{C}^N\mathbf{C} + \mathbf{C}\mathbf{C}^N). \tag{4.4}$$

On applying (4.2) for $N$, (4.4) becomes

$$\mathbf{C}^{N+1} = \tfrac{1}{2}\sum_{j=0}^{m}\sum_{i=0}^{m} u_{Ni}c_j(\mathbf{B}_i\mathbf{B}_j + \mathbf{B}_j\mathbf{B}_i). \tag{4.5}$$

Hence substituting for $\mathbf{B}_i\mathbf{B}_j + \mathbf{B}_j\mathbf{B}_i$ from (4.1),

$$\mathbf{C}^{N+1} = \sum_{t=0}^{m}\left\{\sum_{i=0}^{m}\sum_{j=0}^{m} u_{Ni}c_jq_{ij}^t\right\}\mathbf{B}_t\,. \tag{4.6}$$

Hence by mathematical induction Lemma 4.1 is proved.

THEOREM 4.1: *If the* $\mathbf{C}$*-matrix of a connected design is* $\mathbf{C} = \sum_0^m c_i\mathbf{B}_i$, *and the matrices* $\mathbf{B}_0, \mathbf{B}_1, \cdots, \mathbf{B}_m$ *are the association matrices of the design satisfying conditions* $\mathbf{B}_0 = \mathbf{I}(v)$ *and* $\mathbf{B}_t\mathbf{B}_s + \mathbf{B}_s\mathbf{B}_t = 2\sum_0^m q_{ts}^i\mathbf{B}_i$, *then the analysis of the design will be identical with that of a* PBIB *design.*

PROOF: From Corollary 3.2.1, and Lemma 4.1, it follows that a solution $\hat{\mathbf{t}} = \mathbf{AQ}$ of $\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}}$ exists such that

$$\mathbf{A} = \sum_{i=0}^{m} e_i\mathbf{B}_i \tag{4.7}$$

and

$$\begin{aligned} \mathbf{I}(v) - \frac{1}{v}\,\mathbf{E}(v, v) &= \mathbf{AC} = \mathbf{CA}, \\ &= \tfrac{1}{2}(\mathbf{AC} + \mathbf{CA}). \end{aligned} \tag{4.8}$$

Simplifying both the sides in terms of $\mathbf{B}_t$'s, we get

$$\mathbf{B}_0 - \frac{1}{v}\sum_{t=0}^{m}\mathbf{B}_t = \sum_{t=0}^{m}\left\{\sum_{j=0}^{m}\sum_{i=0}^{m} c_ie_jq_{ij}^t\right\}\mathbf{B}_t\,. \tag{4.9}$$

Hence, on equating the coefficients of the matrices $\mathbf{B}_t$ on both sides of the equation, the $e_i$'s are given by a solution of the equations

$$\begin{aligned} \sum_{i=0}^{m}\sum_{j=0}^{m} q_{ij}^t c_ie_j &= 1 - \frac{1}{v}, & \text{if } t = 0; \\ &= -\frac{1}{v}, & \text{if } t = 1, 2, \cdots, m \end{aligned} \tag{4.10}$$

On comparing equations (4.10) and (3.13), they are seen to be identical except for a change of notation. This implies that one can obtain exactly the same analysis as that of a PBIB design (Bose and Nair [2]), even if the condition (3.5) is replaced by the less stringent condition (4.1).

The combinatorial implication of the condition (4.1) is the following: If two treatments are $i$th associates, then the number of treatments common between the $j$th associates of the first and the $k$th associates of the second, plus the number of treatments common between the $k$th associates of the first and the $j$th associates of the second, is equal to $2q_{jk}^{i}$, and is the same for all the pairs of treatments which are $i$th associates.

Hence the above condition can replace the condition (iiic) of the definition of a PBIB design given by Bose and Shimamoto [3], and the analysis of the design will remain the same. In the case of two associate classes the two conditions are equivalent, but in general they are not.

*Example* 4.1: Consider the following design with parameters: $v = 6$, $b = 9$, $r = 3$, $k = 2$, $m = 4$, $n_1 = n_2 = n_3 = 1$, $n_4 = 2$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = \lambda_4 = 0$. The plan of the design is given in the Table 4.1 and the association scheme in Table 4.2.

Now consider the treatments 1 and 3. The number of treatments common between the 1st associates of 1 and the 2nd associates of 3 is one, whereas there is no treatment common between the 1st associates of 3 and the 2nd associates of 1. Hence it is clear that this design is not a PBIB as defined by [3], but it can be verified that the design satisfies the condition given in (4.1) and that some of the $q_{jk}$ are

$$q_{12}^{4} = \tfrac{1}{2} = q_{23}^{4} = q_{31}^{4}\,. \tag{4.11}$$

One observes that the above example is obtained by taking two $X$-replications and one $Y$-replication of a $3 \times 2$ simple rectangular lattice design (Harshberger [5]). A similar result will be obtained for any design formed by taking $r_1$ $X$-replications and $r_2$ $Y$-replications ($r_1 \neq r_2$) of a $p(p-1)$ simple rectangular lattice design; but, in general, when $p > 3$, there will be five associate classes.

**5. Further generalisation.** From the foregoing arguments, we can see that an analysis almost similar to that of a PBIB can be derived from only the assumptions that association matrices satisfy the condition (4.1) and that the **C**-matrix and $\mathbf{I}(v)$ are linear combinations of the association matrices. Hence, instead of taking $\mathbf{B}_0 = \mathbf{I}(v)$, we can think of some association matrices yielding $\mathbf{I}(v)$ as their linear combination. This will lead to partitioning treatments into several groups and finally, to the following definition:

*Definition* 5.1: In an incomplete block design, partial balance over intra- and inter-group treatment comparisons will be achieved, if the following conditions are satisfied:

(i) The experimental material is divided into $b$ blocks of $k$ plots each, different treatments being applied to the units in the same block.

TABLE 4.1.
*Plan of the design*

| Replication | 1 | | | 2 | | | 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Block | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Treatments | 1<br>2 | 3<br>4 | 5<br>6 | 1<br>2 | 3<br>4 | 5<br>6 | 1<br>6 | 2<br>3 | 4<br>5 |

TABLE 4.2.
*Association scheme*

| Treatment | Associates | | | |
|---|---|---|---|---|
| | 1st | 2nd | 3rd | 4th |
| 1 | 2 | 6 | 4 | 3, 5 |
| 2 | 1 | 3 | 5 | 4, 6 |
| 3 | 4 | 2 | 6 | 1, 5 |
| 4 | 3 | 5 | 1 | 2, 6 |
| 5 | 6 | 4 | 2 | 1, 3 |
| 6 | 5 | 1 | 3 | 2, 4 |

(ii) There are $v$ treatments divided into $h$ groups of $n_1, n_2, \cdots, n_h$ treatments respectively; the treatments of the $i$th group occur in exactly $r_i$ blocks.

(iii) There can be established relations of association between any two treatments satisfying the following requirements:

(a) A treatment of the $i$th group and a treatment of the $j$th group are either $ij$:1th, $ij$:2th, $\cdots$, or $ij:m_{ij}$th associates $(i, j = 1, 2, \cdots, h)$; $ij:t$th associates are the same as $ji:t$th associates.

(b) Each treatment of the $i$th group has exactly $n_{ij:t}ij:t$th associates $(j = 1, 2, \cdots, h, t = 1, 2, \cdots, m_{ij})$ and has zero $lk:t$th associates $(l \neq i, k \neq i)$.

(c) Given any two treatments which are the $ij:t$th associates, the number of treatments common to the $i_1j_1:t_1$th associates of the first and $i_2j_2:t_2$th associates of the second plus the number of treatments common to the $i_2j_2:t_2$th associates of the first and $i_1j_1:t_1$th associates of the second is $2\, q_{ij:t}(i_1j_1:t_1, i_2j_2:t_2)$ and is independent of the pair of the treatments with which we start.

(iv) Two treatments which are $ij:t$th associates occur together in exactly $\lambda_{ij:t}$ blocks.

Because of the treatment groupings the condition (iiic) of Definition 5.1 can be expressed as follows:

(d) Given any two treatments which are the $ij:t$the associates $(i \neq j)$, the first belonging to the $i$th group and the second belonging to the $j$th

group, the number of treatments common to $ik:t_1$th associates of the first and $jk:t_2$th associates of the second is equal to $2\ q_{ij:t}(ik:t_1\,,\ jk:t_2)$ and is independent of the pair of treatments with which we start. Also given any two treatments which are the $ii:t$th associates, the number of treatments common to the $ik:t_1$ associates of the first and $ik:t_2$th associates of the second plus the number of treatments common to the $ik:t_2$th associates of the first and $ik:t_1$th associates of the second is equal to $2\ q_{ii:t}(ik:t_1\,,\ ik:t_2)$ and is independent of the pair of treatments with which we start.

In these designs the total number of associate classes '$m$' is given by

$$m = \sum_{i \geq j = 1}^{h} m_{ij}\,. \tag{5.5}$$

The relations between the parameters are

$$\sum_{i=1}^{h} n_i = v,$$

$$n_j = \sum_{t=0}^{m_{jj}} n_{jj:t} = \sum_{t=1}^{m_{ij}} n_{ij:t}\,, \qquad i \neq j; \tag{5.6}$$

$$\sum_{l=1}^{m_{ik}} q_{ii:t}(ik:t_1\,,\ ik:l) = n_{ik:t_1}\,,$$

$$2\sum_{l=1}^{m_{jk}} q_{ij:t}(ik:t_1\,,jk:l) = n_{ik:t_1}\,, \qquad \text{if } i \neq j.$$

$$n_{ij:t}q_{ij:t}(ik:l,\ jk:t_1) = n_{ik:l}q_{ik:l}(ij:t,\ jk:t_1),$$
$$\text{if } i \neq j,\ i \neq k.$$

If $\mathbf{B}_{ij:t}$ denotes the association matrix corresponding to the $ij:t$th associate class, then

$$\mathbf{B}_{i_1j_1:t_1}\mathbf{B}_{i_2j_2:t_2} + \mathbf{B}_{i_2j_2:t_2}\mathbf{B}_{i_1j_1:t_1} = 2\sum{}^{*}\ q_{ij:t}(i_1j_1:t_1\,,\ i_2j_2:t_2)\mathbf{B}_{ij:t}\,, \tag{5.7}$$

where $\sum^*$ denotes the summation over all the possible values of $ij:t$.

Also, the $\mathbf{C}$ matrix can be written in the form

$$\mathbf{C} = \sum{}^{*}\ c_{ij:t}\mathbf{B}_{ij:t}\,, \tag{5.8}$$

where

$$\begin{aligned} c_{ij:t} &= r_i(k-1)/k, && \text{if } i = j \text{ and } t = 0; \\ &= -\,\lambda_{ij:t}/k, && \text{otherwise.} \end{aligned} \tag{5.9}$$

Hence, by Lemma 4.1 and Corollary 3.2.1, the solution of the normal equations is given by $\hat{\mathbf{t}} = \mathbf{AQ}$ where the matrix $\mathbf{A}$ is of the form

$$\mathbf{A} = \sum{}^{*}\ d_{ij:t}\mathbf{B}_{ij:t}\,, \tag{5.10}$$

and the constants $d_{ij:t}$ are given by a solution of the equations

$$(5.11) \qquad \sum{}' q_{ij:t}(i_1j_1:t_1, i_2j_2:t_2)c_{i_1j_1:t_1} d_{i_2j_2:t_2} = 1 - v^{-1}, \text{ if } i = j \text{ and } t = 0;$$
$$= -v^{-1} \text{ otherwise,}$$

where $\sum'$ represents the summation over all the values of $i_1j_1:t_1$ and $i_2j_2:t_2$ Now, from Lemmas 2.1 and 2.2, it can be assumed that a solution, such that **A** is orthogonal to the vector $\mathbf{E}(v, 1)$, exists, and then

$$(5.10) \qquad d_{ii:0} + \sum_{j=1}^{h} \sum_{t=1}^{m_{ij}} n_{ij:t}\, d_{ij:t} = 0.$$

Hence, using $h$ equations of (5.10), $(m + h)$ equations of (5.11) can be reduced to $m$ equations in $m$ unknowns. So it seems that the analysis of the designs given in Definition 5.1 is similar to that of a PBIB design with $m$ associate classes.

In general, these designs involve a large number of associate classes and consequently their analysis is complicated. The minimum number of classes $m$ is 3, when $h = 2$; the analysis for this design is given by Nair and Rao [9].

Another simple case is the one for which $m_{ij} = 1$ and $\lambda_{ij:1} = \lambda$ for all $i \neq j$. In this case the inverse of $\mathbf{C} + (\lambda/k)\mathbf{E}(v, v)$ can be obtained by working out the inverses of $h$ diagonal sub-matrices. Further, if $m_{ii} = 1$ or 2, the computational work will be reduced considerably.

**6. Acknowledgement.** The author is grateful to Professor M. C. Chakrabarti for his help and guidance.

## REFERENCES


[1] R. C. Bose and Dale M. Mesner, "On linear associative algebra corresponding to association schemes of partially balanced designs," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 21–38.

[2] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939), pp. 337–372.

[3] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[4] Boyd Harshberger, "Preliminary report on rectangular lattices," *Biometrics*, Vol. 2 (1946), pp. 115–119.

[5] C. C. MacDuffee, *The Theory of Matrices*, Chelsea Publishing Co., New York, 1946.

[6] K. R. Nair, "Rectangular lattices and partially balanced incomplete block designs," *Biometrics*, Vol. 7 (1951), pp. 145–154.

[7] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block designs," *Science and Culture*, Vol. 7 (1942), pp. 568–569.

[8] K. R. Nair and C. R. Rao, "Incomplete block designs for experiments involving several groups of varieties," *Science and Culture*, Vol. 7 (1942), pp. 615–616.

# THE NON-EXISTENCE OF CERTAIN PBIB DESIGNS

By Manohar Narhar Vartak

*University of Bombay*

**1. Introduction.** Let $N$ be a Partially Balanced Incomplete Block (PBIB) design, (cf. Bose and Shimamoto, [1]), with three associate classes and with parameters

$$v, b, r, k, n_i, \lambda_i, p^i_{ju}; \qquad (i, j, u = 1, 2, 3). \tag{1.1}$$

These parameters are not all independent but they are connected by the equations

$$\begin{gathered} bk = vr; \qquad \sum_{i=1}^{3} n_i = v - 1; \qquad \sum_{i=1}^{3} n_i\lambda_i = r(k-1); \\ p^i_{ju} = p^i_{uj}; \qquad n_i\, p^i_{ju} = n_j\, p^j_{iu} = n_u\, p^u_{ij}; \\ \sum_{u=1}^{3} p^i_{ju} = n_j - \delta_{ij} \qquad (i, j, u = 1, 2, 3); \end{gathered} \tag{1.2}$$

where $\delta_{ij} = 0$ or 1 according as $i \neq j$ or $i = j$ respectively. Additional relations among the parameters (1.1) can be derived if the association scheme of the $v$ treatments of $N$ is completely known. Suppose, for example, that the association scheme of the given design $N$ is of the rectangular type; that is, let us suppose that

$$v = v_1v_2 \qquad (v_1, v_2 \geqq 2), \tag{1.3}$$

and that the treatments $\theta_{ij}(i = 1, 2, \cdots, v_1; j = 1, 2, \cdots, v_2)$ of the design $N$ can be arranged in the form of a $v_1 \times v_2$ rectangle

$$\begin{matrix} \theta_{11}, \theta_{12}, \cdots, \theta_{1v_2} \\ \theta_{21}, \theta_{22}, \cdots, \theta_{2v_2} \\ \cdots\cdots\cdots\cdots \\ \theta_{v_11}, \theta_{v_12}, \cdots, \theta_{v_1v_2} \end{matrix} \tag{1.4}$$

so that the first associates of any treatment $\theta_{ij}$ are the other $v_2 - 1$ treatments in the $i$th row; its second associates are the other $v_1 - 1$ treatments in the $j$th column and the remaining $(v_1 - 1)(v_2 - 1)$ treatments are its third associates. For the design $N$ with the association scheme (1.4) it then follows that the matrices $(p^i_{ju})$ are given by

Received July 22, 1957; revised July 7, 1959.

$$(p^1_{ju}) = \begin{bmatrix} v_2 - 2 & 0 & 0 \\ 0 & 0 & v_1 - 1 \\ 0 & v_1 - 1 & (v_1 - 1)(v_2 - 2) \end{bmatrix};$$

$$(1.5) \qquad (p^2_{ju}) = \begin{bmatrix} 0 & 0 & v_2 - 1 \\ 0 & v_1 - 2 & 0 \\ v_2 - 1 & 0 & (v_1 - 2)(v_2 - 1) \end{bmatrix};$$

$$(p^3_{ju}) = \begin{bmatrix} 0 & 1 & v_2 - 2 \\ 1 & 0 & v_1 - 2 \\ v_2 - 2 & v_1 - 2 & (v_1 - 2)(v_2 - 2) \end{bmatrix}.$$

The relevant additional relations among the parameters (1.1) are, in this case,

$$(1.6) \qquad p^1_{11} = n_1 - 1 = v_2 - 2; \quad p^2_{22} = n_2 - 1 = v_1 - 2; \quad n_3 = n_1 n_2 .$$

The parameters $r$, $k$, $\lambda_1$, $\lambda_2$ and $\lambda_3$ are related to $v_1$ and $v_2$ through the equation

$$(1.7) \qquad r(k - 1) = (v_2 - 1)\lambda_1 + (v_1 - 1)\lambda_2 + (v_1 - 1)(v_1 - 1)\lambda_3$$

which, in fact, is one of the equations in (1.2) rewritten in the light of (1.6).

In this paper we shall be concerned with PBIB designs with three associate classes whose parameters satisfy the conditions (1.3), (1.5), (1.6) and (1.7) in addition to (1.2). We shall call the series of these designs the series A. A design belonging to the series A will be said to be symmetric if

$$(1.8) \qquad v = b, \quad \text{and consequently,} \quad r = k.$$

It may be noted that the series A includes all PBIB designs with three associate classes which are the Kronecker product of two BIB designs (cf. Vartak [2]).

In the next section we shall show that the conditions (1.2) and (1.6) uniquely characterise the association scheme (1.4). We shall then obtain an expression for the matrix $NN'$ for any design belonging to the series A where $N$ is the incidence matrix of the given design and $N'$ is the transpose of $N$. In Section 3 we shall calculate the characteristic roots and the determinant $| NN' |$ of the matrix $NN'$. We shall also calculate there the Hasse-Minkowski invariants, $c_p(NN')$, for the matrix $NN'$ of any design belonging to the series A.

Some non-existence theorems together with illustrations are given in Section 4. These theorems are direct consequences of the results obtained in Sections 2 and 3, and consist of extensions of the results of Schützenberger [3] and Shrikhande [4] for symmetrical BIB designs, applicable to the designs of series A.

**2. The uniqueness of the rectangular association scheme.** We shall first prove the following theorem on the uniqueness:

THEOREM 2.1: *If the parameters of a* PBIB *design N with three associate classes satisfy the conditions* (1.2) *and* (1.6), *i.e., if the design belongs to the series* A,

*then the association scheme for its treatments is uniquely determined and is of the rectangular type* (1.4).

PROOF: From (1.2) and (1.6) we have, first of all,

$$v = n_1 + n_2 + n_3 + 1 = (v_2 - 1) + (v_1 - 1) + (v_1 - 1)(v_2 - 1) + 1 = v_1 v_2,$$

which is the same as (1.3). Also from (1.2) and (1.6) it follows that the matrices $(p^i_{ju})$ are as given in (1.5).

Let $\phi$ and $\theta$ be any two treatments of $N$ which are first associates. Let $\phi_{11}, \cdots, \phi_{1n_1}$ be the $n_1$ first associates of $\phi$ and $\theta_{11}, \cdots, \theta_{1n_1}$ be the $n_1$ first associates of $\theta$. Then $\phi$ is one of the $\theta_{1i}$'s and $\theta$ is one of the $\phi_{1i}$'s $(i = 1, 2, \cdots, n_1)$. Let us say, for the sake of definiteness, that $\phi_{11} \equiv \theta$ and $\theta_{11} \equiv \phi$. Now, since by (1.6), $p^1_{11} = n_1 - 1 = v_2 - 2$, it follows that the sets $\phi_{1i}$ and $\theta_{1i}$ have exactly $v_2 - 2 = n_1 - 1$ treatments in common. From this and the earlier identifications $\phi_{11} \equiv \theta$ and $\theta_{11} \equiv \phi$, it follows that the sets $\phi_{1j}$ and $\theta_{1j}$ $(j = 2, 3, \cdots, n_1)$ are identical, i.e., consist of the same treatments. This means that any two treatments in the set $\{\phi, \theta, \theta_{12}, \cdots, \theta_{1n_1}\}$ $(n_1 = v_2 - 1)$, of $v_2$ treatments, are first associates and that the remaining $v_2 - 2$ treatments are first associates of each of them. This implies that the relation of being first associates is symmetric as well as transitive for all treatments of the design $N$. From this it follows that the $v = v_1 v_2$ treatments of the design $N$ fall into $v_1$ groups of $v_2$ treatments each, such that the relation of being first associates is symmetric as well as transitive for the treatments of any of the $v_1$ groups. It is, therefore, convenient to designate these groups by

$$\begin{array}{c} (\theta_{11}, \ \theta_{12}, \ \cdots, \ \theta_{1v_2}) \\ (\theta_{21}, \ \theta_{22}, \ \cdots, \ \theta_{2v_2}) \\ \cdot \ \ \cdot \ \ \cdot \ \ \cdot \ \ \cdot \ \ \cdot \ \ \cdot \\ (\theta_{v_1 1}, \theta_{v_1 2}, \cdots, \theta_{v_1 v_2}). \end{array} \tag{2.1}$$

The property satisfied by any of these groups is that the first associates of any treatment in the group are the remaining treatments in the same group.

Next, suppose that the second group in (2.1) contains two treatments $\theta_{2j}$ and $\theta_{2k}$ which are second associates of $\theta_{11}$. This will mean that $p^2_{21} \geqq 1$, which contradicts the result $p^2_{21} = 0$ obtained earlier and referred to in (1.5). This implies that the second, and in general any of the $v_1 - 1$ groups after the first, cannot contain more than one second associate of $\theta_{11}$. But $\theta_{11}$ has exactly $n_2 = v_1 - 1$ second associates so that the 2nd, 3rd, $\cdots$, $v_1$th group in (2.1) must each contain one and only one second associate of $\theta_{11}$. The same holds for each of $\theta_{12}, \cdots, \theta_{1v_2}$. In general, therefore, the $i$th group contains one and only one second associate of $\theta_{jk}$ when $j \neq i$. Without any loss of generality, we can assume that $\theta_{2i}, \theta_{3i}, \cdots, \theta_{v_1 i}$ are the $n_2 = v_1 - 1$ second associates of $\theta_{1i}$.

Further, we have $p^2_{22} = n_2 - 1 = v_1 - 2$, which, by the same type of argument as before, implies that the treatments $\theta_{1i}, \theta_{2i}, \cdots, \theta_{v_1 i}$ are such that the relation of being second associates is symmetric as well as transitive for them. The $v = v_1 v_2$

treatments of $N$, therefore, can be conveniently divided into $v_2$ groups of $v_1$ treatments each, such that the relation of being second associates is symmetric as well as transitive for the treatment of any group.

The two modes of classification of the treatments of $N$ for the relation of first and second association can be superimposed by writing the treatments in the form of a rectangular array (1.4).

The third associates of any treatment $\theta_{ij}$ are, then, by exclusion, the $n_3 = (v_2 - 1)(v_1 - 1) = n_1 n_2$ treatments $\theta_{kl}$ in the array, where $k \neq i$ and $l \neq j$. The relation of association for the treatments of the design $N$ can thus be described with the help of the association scheme (1.4), where the treatments occurring in the same row as $\theta_{ij}$ are its first associates, those occurring in the same column as $\theta_{ij}$ are its second associates, and the others are its third associates. In other words the association scheme is uniquely determined and is of the rectangular type.

This proves the theorem.

With the help of the association scheme (1.4), we can write down the matrix $NN'$ of the design $N$ belonging to the series A in a very convenient form. Let the rows of $N$ correspond to the treatments $\theta_{11}, \theta_{12}, \cdots, \theta_{1v_2}, \theta_{21}, \cdots \theta_{2v_2}, \cdots, \theta_{v_1 1}, \cdots, \theta_{v_1 v_2}$ respectively, in this order. Then the matrix $NN'$ is seen to have the following structure:

$$NN' = \begin{bmatrix} A & B & \cdots & B \\ B & A & \cdots & B \\ \cdots & \cdots & \cdots & \cdots \\ B & B & \cdots & A \end{bmatrix} \tag{2.2}$$

where $A$ is a $v_2 \times v_2$ square matrix given by

$$A = (r - \lambda_1) I_{v_2} + \lambda_1 E_{v_2} \tag{2.3}$$

and $B$ is a $v_2 \times v_2$ square matrix given by

$$B = (\lambda_2 - \lambda_3) I_{v_2} + \lambda_3 E_{v_2}, \tag{2.4}$$

$I_{v_2}$ being the identity matrix of order $v_2$ and $E_{v_2}$ a square matrix of order $v_2$ with all elements equal to 1. Also the matrix $NN'$, as written in (2.2), has $v_1$ rows and $v_1$ columns. The same result can be summarized in the form of the following theorem:

THEOREM 2.2: *The matrix $NN'$ for a design $N$ belonging to the series* A *is given by*

$$NN' = I_{v_1} \times (A - B) + E_{v_1} \times B \tag{2.5}$$

*where '$\times$' denotes the Kronecker product of matrices and $A$ and $B$ are as defined in* (2.3) *and* (2.4).

**3. Characteristic roots, determinant and the Hasse-Minkowski invariants of $NN'$.** Let $D_{v_2}$ be the $v_2 \times v_2$ square matrix given by

$$(3.1)\qquad D_{v_2} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 1 & -2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & -(v_2 - 1) \end{bmatrix}$$

It should be observed that the matrix $D_{v_2}$ is a modified Helmertz matrix. Moreover, the determinant $|D_{v_2}|$ of $D_{v_2}$ is clearly

$$(3.2)\qquad |D_{v_2}| = (-)^{v_2-1}\{v_2!\},$$

so that $D_{v_2}$ is non-singular. In fact $D_{v_2}$ is a semi-orthogonal matrix in the sense that

$$(3.3)\qquad D_{v_2}D'_{v_2} = \text{diag}\{v_2, 1.2, 2.3, \cdots, (v_2 - 1)v_2\}$$

where $\text{diag}\{a_1, a_2, \cdots, a_m\}$ is a diagonal matrix of order $m$ whose diagonal elements are $a_1, a_2, \cdots, a_m$ and off-diagonal elements are all zero. It is easy to verify that the matrix $D_{v_2}$ reduces both $A$ and $B$ to diagonal forms. Thus

$$(3.4)\qquad D_{v_2}AD'_{v_2} = \text{diag}\{v_2[r + (v_2 - 1)\lambda_1], 1.2\,(r - \lambda_1), 2.3\,(r - \lambda_1), \cdots, (v_2 - 1)v_2(r - \lambda_1)\}$$

and

$$(3.5)\qquad D_{v_2}BD'_{v_2} = \text{diag}\{v_2[\lambda_2 + (v_2 - 1)\,\lambda_3], 1.2(\lambda_2 - \lambda_3), \cdots, (v_2 - 1)v_2(\lambda_2 - \lambda_3)\}.$$

It may be noted that, since the elements of $D_{v_2}$ are all integral, the equations (3.4) and (3.5) can be interpreted to mean that $A$ and $B$ are both rationally equivalent to the diagonal forms exhibited on the right sides of (3.4) and (3.5).

Now consider the matrix

$$(3.6)\qquad H = \begin{bmatrix} D_{v_2} & D_{v_2} & D_{v_2} & \cdots & D_{v_2} \\ D_{v_2} & -D_{v_2} & 0 & \cdots & 0 \\ D_{v_2} & D_{v_2} & -2D_{v_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ D_{v_2} & D_{v_2} & D_{v_2} & \cdots & -(v_1 - 1)D_{v_2} \end{bmatrix}$$

where $D_{v_2}$ is the matrix given by (3.1) and $H$, as written above, has $v_1$ rows and $v_1$ columns, every 0 in (3.6) being a square null matrix of order $v_2 \times v_2$. It may be noted that the matrix $H$ is the Kronecker product $D_{v_1} \times D_{v_2}$ and hence the determinant $|H|$ of $H$ is given by

$$(3.7)\qquad |H| = |D_{v_1} \times D_{v_2}| = |D_{v_1}|^{v_2} \cdot |D_{v_2}|^{v_1} = (-)^{v_1+v_2}(v_1!)^{v_2}(v_2!)^{v_1}.$$

The characteristic roots of $NN'$ of (2.2) are the roots of the determinantal equation in $\theta$:

$$(3.8)\qquad |NN' - \theta I_v| = 0$$

where $I_v (v = v_1 v_2)$ is the identity matrix of order $v$.

From (2.5), we can write this in the form

$$(3.9) \qquad | I_{v_1} \times \{(A - \theta I_{v_2}) - B\} + E_{v_1} \times B | = 0$$

However, it is easy to verify that

$$\begin{aligned}(3.10) \quad H\{NN' - \theta I_v\} H' &= H\{I_{v_1} \times [(A - \theta I_{v_2}) - B] + E_{v_1} \times B\} H' \\ &= \mathrm{diag}\{v_1 D_{v_2}[(A - \theta I_{v_2}) + (v_1 - 1)B]D'_{v_2}, 1.2\, D_{v_2}[(A - \theta I_{v_2}) - B] \\ &\qquad \cdot D'_{v_2}, \cdots, (v_1 - 1)v_1 D_{v_2}[(A - \theta I_{v_2}) - B]D'_{v_2}\}\end{aligned}$$

and since $D_{v_2}AD'_{v_2}$, $D_{v_2}BD'_{v_2}$ and $D_{v_2}D'_{v_2}$ are themselves diagonal matrices, so are $D_{v_2}\{(A - \theta I_{v_2}) - B\} D_{v_2}$ and $D_{v_2}\{(A - \theta I_{v_2}) + (v_1 - 1)B\}D'_{v_2}$. Hence (3.10) reduces completely to a diagonal matrix. Writing

$$(3.11) \qquad \begin{aligned}\theta_0 &= rk = r + (v_2 - 1)\lambda_1 + (v_1 - 1)\lambda_2 + (v_1 - 1)(v_2 - 1)\lambda_3, \\ \theta_1 &= r - \lambda_1 + (v_1 - 1)(\lambda_2 - \lambda_3), \\ \theta_2 &= r - \lambda_2 + (v_2 - 1)(\lambda_1 - \lambda_3), \\ \theta_3 &= r - \lambda_1 - \lambda_2 + \lambda_3,\end{aligned}$$

we find that (3.10) reduces to

$$(3.12) \qquad \begin{aligned}&H\{NN' - \theta I_v\}H' \\ &= \mathrm{diag}\{v_1 v_2(\theta_0 - \theta), v_1\, 1.2(\theta_1 - \theta), \cdots, v_1(v_2 - 1)v_2(\theta_1 - \theta), \\ &\qquad 1.2\, v_2(\theta_2 - \theta), 1.2 \cdot 1.2(\theta_3 - \theta), \cdots, 1.2(v_2 - 1)v_2(\theta_3 - \theta), \\ &\qquad \cdots \\ &\qquad (v_1 - 1)v_1 v_2(\theta_2 - \theta), (v_1 - 1)v_1\, 1.2(\theta_3 - \theta), \cdots, \\ &\qquad\qquad (v_1 - 1)v_1(v_2 - 1)v_2(\theta_3 - \theta)\}.\end{aligned}$$

Hence, taking the determinants of both sides, we get

$$(3.13) \qquad | NN' - \theta I_v | = (\theta_0 - \theta)(\theta_1 - \theta)^{v_2 - 1}(\theta_2 - \theta)^{v_1 - 1}(\theta_3 - \theta)^{(v_1 - 1)(v_2 - 1)}.$$

Also the determinant $| NN' |$ of the matrix $NN'$ is the product of its characteristic roots. Hence from (3.13) and (3.11) we get the following theorem:

THEOREM 3.1:

(a) *The characteristic roots of the matrix $NN'$ of the design $N$ of the series A are $\theta_0$, $\theta_1$, $\theta_2$, $\theta_3$ given by* (3.11) *and their respective multiplicities are*

$$(3.14) \qquad \alpha_0 = 1, \quad \alpha_1 = v_2 - 1 = n_1, \quad \alpha_2 = v_1 - 1 = n_2, \quad \alpha_3 = (v_1 - 1)(v_2 - 1) = n_3.$$

(b) *The determinant $| NN' |$ of the matrix $NN'$ of the design $N$ is given by*

$$| NN' | = \theta_0 \theta_1^{v_2 - 1} \theta_2^{v_1 - 1} \theta_3^{(v_1 - 1)(v_2 - 1)}$$

$$(3.15) \quad = rk\{r - \lambda_1 + (v_1 - 1)(\lambda_2 - \lambda_3)\}^{v_2 - 1}\{r - \lambda_2 + (v_2 - 1)(\lambda_1 - \lambda_3)\}^{v_1 - 1} \cdot \{r - \lambda_1 - \lambda_2 + \lambda_3\}^{(v_1 - 1)(v_2 - 1)}.$$

To derive an expression for the Hasse-Minkowski invariant $c_p(NN')$ of the matrix $NN'$, we note that, from (3.10),

$$HNN'H' = \text{diag}\{v_1 D_{v_2}[A + (v_1 - 1)B]D'_{v_2}, 1.2\, D_{v_2}(A - B)D'_{v_2}, 2.3\, D_{v_2}(A - B)D'_{v_2}, \cdots, (v_1 - 1)v_1 D_{v_2}(A - B)D'_{v_2}\}.$$

This can be further written as the direct sum of the matrix

$$v_1 D_{v_2}[A + (v_1 - 1)B]D'_{v_2}$$

and the Kronecker product

$$\text{diag}\{1.2, 2.3, \cdots, v_1(v_1 - 1)\} \times \{D_{v_2}(A - B)D'_{v_2}\}.$$

That is, we can write

$$(3.16) \quad HNN'H' = v_1 D_{v_2}\{A + (v_1 - 1)B\}D'_{v_2} \dotplus \text{diag}\{1.2, 2.3, \cdots, v_1(v_1 - 1)\} \times \{D_{v_2}(A - B)D'_{v_2}\},$$

where $\dotplus$ denotes the direct sum.

We now make use of the following results for the $c_p$ invariants of the direct sum and the Kronecker product of matrices:

If $P$ and $Q$ are symmetric matrices with rational elements whose $c_p$ invariants are defined and if

$$U = P \dotplus Q \text{ and } V = P \times Q$$

then

$$(3.17) \quad c_p(U) = (-1, -1)_p c_p(P) c_p(Q)(|P|, |Q|)_p,$$

and

$$(3.18) \quad c_p(V) = (-1, -1)_p^{m+n-1} \{c_p(P)\}^n \{c_p(Q)\}^m (|P|, -1)_p^{\frac{n(n-1)}{2}} (|Q|, -1)_p^{\frac{m(m-1)}{2}} (|P|, |Q|)_p^{mn-1}$$

where $m$ and $n$ are the orders of $P$ and $Q$ respectively, (cf. [5] and [6] respectively).

Further we know that if $\lambda$ is a non-zero rational number and $B$ is an $n \times n$ matrix whose Hasse-Minkowski invariants are defined, then

$$(3.19) \quad c_p(\lambda B) = c_p(B)(\lambda, -1)_p^{\frac{n(n+1)}{2}} (\lambda, |B|)_p^{n-1}$$

where $|B|$ is the determinant of $B$.

It should be noted that $HNN'H'$ of (3.16) is rationally equivalent to $NN'$ and is a diagonal matrix.

We are now in a position to prove the following theorem:

THEOREM 3.2: *The Hasse-Minkowski invariant $c_p(NN')$ of the matrix $NN'$ for the design $N$ of the series* A *is given by*

$$
\begin{aligned}
c_p(NN') = {} & (-1,-1)_p(\theta_0,-v)_p(v_1\theta_0,\theta_1)_p^{v_2-1}(v_2\theta_0,\theta_2)_p^{v_1-1}(\theta_0,\theta_3)_p^{(v_1-1)(v_2-1)} \\
& \times \{(\theta_1,\theta_2)_p(\theta_2,\theta_3)_p(\theta_3,\theta_1)_p\}^{(v_1-1)(v_2-1)} \\
& \times (\theta_1,-1)_p^{\frac{v_2(v_2-1)}{2}}(\theta_2,-1)_p^{\frac{v_1(v_1-1)}{2}}(\theta_3,-1)_p^{\frac{(v_1-1)(v_2-1)(v_1+v_2-2)}{2}} \\
& \times (\theta_1,v_2)_p(\theta_2,v_1)_p(\theta_3,v_1)_p^{v_2-1}(\theta_3,v_2)_p^{v_1-1},
\end{aligned} \tag{3.20}
$$

*if the characteristic roots $\theta_0$, $\theta_1$, $\theta_2$, $\theta_3$ of the matrix $NN'$, given by* (3.11), *are all non-zero.*

PROOF: Observe, in the first place, that

$$D_{v_2}\{A+(v_1-1)B\}D'_{v_2} = \mathrm{diag}\{v_2\theta_0, 1.2\,\theta_1, 2.3\,\theta_1, \cdots, v_2(v_2-1)\theta_1\}, \tag{3.21}$$

and that

$$D_{v_2}(A-B)D'_{v_2} = \mathrm{diag}\{v_2\theta_2, 1.2\,\theta_3, 2.3\,\theta_3, \cdots, v_2(v_2-1)\theta_3\}. \tag{3.22}$$

Hence, when the characteristic roots $\theta_0$, $\theta_1$, $\theta_2$, $\theta_3$ are all non-zero, from (3.16) we find that all the leading principal minor determinants of the rationally equivalent diagonal form of $NN'$ are different from zero; so that the Hasse-Minkowski invariants of this diagonal form and consequently that of the matrix $NN'$ are defined.

A little algebra shows that

$$c_p\{\mathrm{diag}(1.2, 2.3, \cdots, v_1(v_1-1))\} = (-1,-1)_p, \tag{3.23}$$

$$
\begin{aligned}
c_p\{D_{v_2}[(A+(v_1-1)B]D'_{v_2}\} = {} & (-1,-1)_p(\theta_0,-v_2)_p \\
& (\theta_0,\theta_1)_p^{v_2-1}(\theta_1,v_2)_p(\theta_1,-1)_p^{\frac{v_2(v_2-1)}{2}},
\end{aligned} \tag{3.24}
$$

$$
\begin{aligned}
c_p\{D_{v_2}(A-B)D'_{v_2}\} = {} & (-1,-1)_p(\theta_2,-v_2)_p \\
& (\theta_2,\theta_3)_p^{v_2-1}(\theta_3,v_2)_p(\theta_3,-1)_p^{\frac{v_2(v_2-1)}{2}}.
\end{aligned} \tag{3.25}
$$

Making use of (3.23), (3.24), (3.25) and (3.17), (3.18) and (3.19), it is possible to obtain (3.20) after a little calculation.

This completes the proof of the theorem.

**4. The non-existence theorems with illustrations.** Let $N$ be a design of the series A characterised by (1.3), (1.5) through (1.7). Let $\chi$ be any characteristic root of $NN'$ for this design. Then there exists a vector $\mathbf{x}$ such that

$$\mathbf{x}'NN'\mathbf{x} = \chi \tag{4.1}$$

which shows that $\chi$ is non-negative. This gives the following theorem:

THEOREM 4.1: *For a design in the series* A *to exist it is necessary that*

$$\theta_1 = r - \lambda_1 + (v_1 - 1)(\lambda_2 - \lambda_3) \geqq 0,$$
$$\theta_2 = r - \lambda_2 + (v_2 - 1)(\lambda_1 - \lambda_3) \geqq 0,$$
$$\theta_3 = r - \lambda_1 - \lambda_2 + \lambda_3 \geqq 0.$$

The following examples illustrate the use of this theorem:

*Example* 4.1: Consider the symmetric ($v = b$ and hence $r = k$) PBIB design of the series A given by

$$v = b = 24, \quad r = k = 8, \quad n_1 = 5, \quad n_2 = 3, \quad n_3 = 15,$$
$$\lambda_1 = 4, \quad \lambda_2 = 7, \quad \lambda_3 = 1;$$
$$(p_{ju}^1) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 3 & 12 \end{bmatrix}, \quad (p_{ju}^2) = \begin{bmatrix} 0 & 0 & 5 \\ 0 & 2 & 0 \\ 5 & 0 & 10 \end{bmatrix}, \quad (p_{ju}^3) = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 2 \\ 4 & 2 & 8 \end{bmatrix}.$$

The characteristic roots of $NN'$ for this design are

$$\theta_0 = 64, \quad \theta_1 = 22, \quad \theta_2 = 16, \quad \theta_3 = -2;$$

and since $\theta_3 < 0$, Theorem 4.1 is contradicted. Hence the above PBIB is impossible.

*Example* 4.2: Consider the PBIB design of the series given by

$$v = 30, \quad b = 20, \quad r = 10, \quad k = 15, \quad n_1 = 4, \quad n_2 = 5, \quad n_3 = 20,$$
$$\lambda_1 = 10, \quad \lambda_2 = 8, \quad \lambda_3 = 3;$$
$$(p_{ju}^1) = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 5 \\ 0 & 5 & 15 \end{bmatrix}, \quad (p_{ju}^2) = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 16 \end{bmatrix}, \quad (p_{ju}^3) = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 4 \\ 3 & 4 & 12 \end{bmatrix}.$$

The characteristic roots of $NN'$ for this design are

$$\theta_0 = 150, \quad \theta_1 = 25, \quad \theta_2 = 30, \quad \theta_3 = -5;$$

and since $\theta_3 < 0$, Theorem 4.1 is contradicted. Hence the above PBIB design is impossible.

*Example* 4.3: Consider the PBIB design of the series A given by

$$v = 30, \quad b = 50, \quad r = 10, \quad k = 6, \quad n_1 = 4, \quad n_2 = 5, \quad n_3 = 20,$$
$$\lambda_1 = 5, \quad \lambda_2 = 6, \quad \lambda_3 = 0;$$
$$(p_{ju}^1) = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 5 \\ 0 & 5 & 15 \end{bmatrix}, \quad (p_{ju}^2) = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 16 \end{bmatrix}, \quad (p_{ju}^3) = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 4 \\ 3 & 4 & 12 \end{bmatrix}.$$

The characteristic roots of $NN'$ for this design are

$$\theta_0 = 60, \quad \theta_1 = 35, \quad \theta_2 = 24, \quad \theta_3 = -1;$$

and since $\theta_3 < 0$, Theorem 4.1 is contradicted. Hence the above PBIB design is impossible.

In the case of a symmetric PBIB design of the series A we have $v = b$ so that the matrix $N$ is a square matrix of order $v = v_1v_2$. The determinant $|NN'|$ of the matrix $NN'$ must therefore be a perfect square when $|N| \neq 0$. This condition can be formulated in the form of the theorem:

THEOREM 4.2: *A necessary condition for the existence of a symmetric* PBIB *design of the series* A *when* $|N| \neq 0$ *is that*

(a) *if* $v_1$ *is even and* $v_2$ *is odd then* $\theta_2 = r - \lambda_2 + (v_2 - 1)(\lambda_1 - \lambda_3)$ *is a perfect square,*

(b) *if* $v_2$ *is even and* $v_1$ *is odd, then* $\theta_1 = r - \lambda_1 + (v_1 - 1)(\lambda_2 - \lambda_3)$ *is a perfect square, and*

(c) *if* $v_1$ *and* $v_2$ *are both even then* $\theta_1\theta_2\theta_3$, $(\theta_3 = r - \lambda_1 - \lambda_2 + \lambda_3)$, *is a perfect square.*

The following examples illustrate the application of this theorem:

*Example* 4.4: Consider the design given by

$$v = b = 66, \quad r = k = 14, \quad n_1 = 2, \quad n_2 = 21, \quad n_3 = 42,$$

$$\lambda_1 = 7, \quad \lambda_2 = 4, \quad \lambda_3 = 2$$

$$(p^1_{ju}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 21 \\ 0 & 21 & 21 \end{bmatrix}, \quad (p^2_{ju}) = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 20 & 0 \\ 2 & 0 & 40 \end{bmatrix}, \quad (p^3_{ju}) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 20 \\ 1 & 20 & 20 \end{bmatrix}.$$

Clearly this design is a symmetric design ($v = b$) from the series A. Since $v_1 = n_2 + 1 = 22$ is an even integer and $v_2 = n_1 + 1 = 3$ is an odd integer and since $\theta_2 = r - \lambda_2 + (v_2 - 1)(\lambda_1 - \lambda_3) = 20$ is not a perfect square, it follows from Theorem 4.2 that the above PBIB design is impossible. It is easy to verify that $|N| \neq 0$.

It may be observed that the parameters of the above PBIB design are obtained by taking the Kronecker product (cf. [2]) of the BIB designs

$$N_1: v_1 = b_1 = 22, \quad r_1 = k_1 = 7, \quad \lambda_1 = 2$$

and

$$N_2: v_2 = b_2 = 3, \quad r_2 = k_2 = 2, \quad \lambda_2 = 1,$$

of which $N_1$ is already known to be non-existent (cf. Shrikhande [4]).

*Example* 4.5: Consider the PBIB design given by

$$v = b = 48, \quad r = k = 10, \quad n_1 = 7, \quad n_2 = 5, \quad n_3 = 35$$

$$\lambda_1 = 5, \quad \lambda_2 = 4, \quad \lambda_3 = 1$$

$$(p^1_{ju}) = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & 5 \\ 0 & 5 & 30 \end{bmatrix}, \quad (p^2_{ju}) = \begin{bmatrix} 0 & 0 & 7 \\ 0 & 4 & 0 \\ 7 & 0 & 28 \end{bmatrix}, \quad (p^3_{ju}) = \begin{bmatrix} 0 & 1 & 6 \\ 1 & 0 & 4 \\ 6 & 4 & 24 \end{bmatrix},$$

which is a symmetric ($b = v$) design from the series A. Here both $v_1$ and $v_2$ are even and the characteristic roots of $NN'$ for this design are $\theta_0 = 100$, $\theta_1 = 20$,

$\theta_2 = 34$, $\theta_3 = 2$. This implies that $|N| \neq 0$. Moreover, $\theta_1\theta_2\theta_3 = 1360$ is not a perfect square. It follows therefore from Theorem 4.2 that the above design is impossible.

The Hasse-Minkowski invariant $c_p(NN')$ obtained in (3.19) gives us another non-existence theorem for the symmetric designs of the series A.

Let $N$ be a symmetric design of the series A with $|N| \neq 0$. Then the matrix $NN' = B$ for this design is obviously rationally equivalent to $I_v$, the identity matrix of order $v = v_1v_2$. Hence $c_p(NN')$ must be $+1$ for all odd primes $p$. If, for any design, $c_p(NN') = -1$ for some odd prime $p$, then that design will be impossible.

We state this result as the following theorem:

THEOREM 4.3: *If $N$ is a symmetrical design of the series A with $|N| \neq 0$, then a necessary condition for the design $N$ to exist is that $c_p(NN') = +1$ for all odd primes $p$.*

The following examples illustrate the use of this theorem:

*Example* 4.6: Consider the PBIB design given by

$$v = b = 87, \quad r = k = 16, \quad n_1 = 28, \quad n_2 = 2, \quad n_3 = 56,$$

$$\lambda_1 = 4, \quad \lambda_2 = 8, \quad \lambda_3 = 2.$$

$$(p^1_{ju}) = \begin{bmatrix} 27 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 54 \end{bmatrix}, \quad (p^2_{ju}) = \begin{bmatrix} 0 & 0 & 28 \\ 0 & 1 & 0 \\ 28 & 0 & 28 \end{bmatrix}, \quad (p^3_{ju}) = \begin{bmatrix} 0 & 1 & 27 \\ 1 & 0 & 1 \\ 27 & 1 & 27 \end{bmatrix}.$$

This is evidently a symmetric design from the series A with $|N| \neq 0$. Further it is easy to verify that $c_p(NN')$ given by (3.19) reduces in this case to $(24, 29)_p$; further, for $p = 3$ this becomes $c_3(NN') = (2, 3)_3 = (2/3) = -1$ where $(a/p)$ is the Legendre symbol of $a$ with respect to the prime $p$. Thus Theorem 4.3 is contradicted and therefore the above design is impossible.

It may be observed that the above design has a set of parameters which could be obtained by taking the Kronecker product of the BIB designs

$$N_1: v_1 = b_1 = 3, \quad r_1 = k_1 = 2, \quad \lambda_1 = 1,$$

and

$$N_2: v_2 = b_2 = 29, \quad r_2 = k_2 = 8, \quad \lambda_2 = 2,$$

of which, $N_2$ is proved to be impossible (cf. Shrikhande [4]).

*Example* 4.7: Consider the PBIB design given by

$$v = b = 63, \quad r = k = 11, \quad n_1 = 8, \quad n_2 = 6, \quad n_3 = 48,$$

$$\lambda_1 = 4, \quad \lambda_2 = 5, \quad \lambda_3 = 1,$$

$$(p^1_{ju}) = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 0 & 6 \\ 0 & 6 & 42 \end{bmatrix}, \quad (p^2_{ju}) = \begin{bmatrix} 0 & 0 & 8 \\ 0 & 5 & 0 \\ 8 & 0 & 40 \end{bmatrix}, \quad (p^3_{ju}) = \begin{bmatrix} 0 & 1 & 7 \\ 1 & 0 & 5 \\ 7 & 5 & 35 \end{bmatrix}.$$

This is obviously a symmetric design from the series A with $|N| \neq 0$. Further it is easy to verify that the Hasse-Minkowski invariant $c_p(NN')$ given by (3.19) reduces in this case to $(30, 7)_p (30, -1)_p$. For $p = 3$ this becomes $c_3(NN') = (2, 3)_3 = (2/3) = -1$, where $(a/p)$ is the Legendre symbol of $a$ with respect to the prime $p$. Thus Theorem 4.3 is contradicted and therefore the above PBIB design is impossible.

**5. Summary and acknowledgement.** Three non-existence theorems are obtained for the PBIB designs with three associate classes and belonging to a certain series called the Series A. The first theorem makes use of the fact that the characteristic roots of the matrix $NN'$ are always non-negative; the second is an extension of Schützenberger's result [3] and the third is an extension of Shrikhande's result [4] for symmetrical BIB designs.

I wish to express my sincere thanks to Professor M. C. Chakrabarti for his kind interest in this work. Also I am indebted to the referee for useful suggestions especially in connection with Theorem 3.2.

## REFERENCES

[1] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *Jour., Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[2] M. N. Vartak, "On an application of Kronecker product of matrices to statistical designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 420–438.

[3] M. P. Schützenberger, "A non-existence theorem for an infinite family of symmetrical block designs," *Ann. Eugenics*, Vol. 14 (1949), pp. 286–287.

[4] S. S. Shrikhande, "The impossibility of certain symmetrical balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 106–111.

[5] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[6] M. N. Vartak, "On the Hasse-Minkowski invariant of the Kronecker product of matrices," *Canadian Jour. Math.*, Vol. 10 (1958), pp. 66–72.

# A NECESSARY CONDITION FOR EXISTENCE OF REGULAR AND SYMMETRICAL EXPERIMENTAL DESIGNS OF TRIANGULAR TYPE, WITH PARTIALLY BALANCED INCOMPLETE BLOCKS[1]

By Junjiro Ogawa

*University of North Carolina*

A necessary condition for the existence of a symmetrical balanced incomplete block (B.I.B.) design in terms of the Hasse-Minkowski $p$-invariant was obtained by S. S. Shrikhande [1]. Similar necessary conditions for regular symmetrical group divisible designs and for regular symmetrical $L_2$ type designs were obtained by R. C. Bose and W. S. Connor [2] and S. S. Shrikhande [3] respectively.

The purpose of this note is to give a necessary condition for the existence of a regular symmetrical partially balanced incomplete block (P.B.I.B.) design of triangular type in terms of the Hasse-Minkowski $p$-invariant.

**1. A necessary theorem and lemmas.** Two symmetric and non-singular matrices $A$ and $B$ of the same order $n$ with rational elements are said to be *rationally congruent* or *congruent in the field of rational numbers*, if there exists a non-singular and rational matrix $C$ of the same order such that

$$C'AC = B, \tag{1.1}$$

where $C'$ stands for the transposed matrix of $C$ [4]. This relation is denoted by the symbol

$$A \sim B. \tag{1.2}$$

By the very definition of the rational congruence, it will be clear that (i) $A \sim A$ (reflexive), (ii) if $A \sim B$, then $B \sim A$ (symmetric), (iii) if $A \sim B$ and $B \sim C$, then $A \sim C$ (transitive), (iv) $A \sim A^{-1}$, and (v) if $A \sim B$, then $A^{-1} \sim B^{-1}$.

*Hasse's Theorem* [4, 5]. The necessary and sufficient conditions for two positive-definite, rational and symmetric matrices $A$ and $B$ of the same order to be rationally congruent are that, in the first place, the square-free parts of the determinants of both matrices are the same, and in the second, the Hasse-Minkowski $p$-invariants of both matrices coincide with each other for all primes $p$ including $p_\infty$.

If we denote the $n$ leading principal minor determinants of $A$ by

$$D_1, D_2, \cdots, D_{n-1}, \qquad D_n = |A|$$

and let $D_0 = 1$, then [4] the Hasse-Minkowski $p$-invariant of $A$ is given by

Received April 6, 1959.

[1] This research was supported by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

$$C_p(A) = (-1, -1)_p \cdot \prod_{i=0}^{n-1} (D_{i+1}, -D_i)_p \tag{1.3}$$

for each prime $p$, where the symbol $(a, b)_p$ denotes the extended Hilbert symbol of norm residue [4, 6], which is defined by

$$(a, b)_p = \begin{cases} +1, & \text{if } ax^2 + by^2 = 1 \text{ has a } p\text{-adic solution} \\ -1, & \text{otherwise.} \end{cases} \tag{1.4}$$

Now we shall list some useful properties of $C_p(A)$ as lemmas.

LEMMA 1.1 [4]: *If $A$ and $B$ are rational and symmetric and if*

$$U = \begin{Vmatrix} A & 0 \\ 0 & B \end{Vmatrix},$$

*then*

$$C_p(U) = (-1, -1)_p(|A|, |B|)_p C_p(A)\, C_p(B). \tag{1.5}$$

LEMMA 1.2 [4]: *For an $n \times n$ diagonal matrix $\Delta_n$, whose $i$, $i$ element is $d$,*

$$C_p(\Delta_n) = (-1, -1)_p(-1, d)_p^{\frac{n(n+1)}{2}}. \tag{1.6}$$

LEMMA 1.3. *For a $(v - 1) \times (v - 1)$ diagonal matrix $U$, whose $i$, $i$ element is*

$$(v - i + 1)\,(v - i),$$

$$C_p(U) = (-1, -1)_p. \tag{1.7}$$

LEMMA 1.4 [4]:

$$C_p(\rho A) = (-1, \rho)_p^{\frac{n(n+1)}{2}} (\rho, |A|)_p^{n-1} C_p(A). \tag{1.8}$$

LEMMA 1.5: *If the $n - 1$ rational vectors*

$$\mathbf{a}_2, \cdots, \mathbf{a}_n$$

*of dimensionality $n$ are linearly independent and are orthogonal to*

$$\mathbf{1}' = (1\ 1 \cdots 1),$$

*then the Gramian of the set, i.e.,*

$$U = \begin{Vmatrix} \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_n' \end{Vmatrix} \cdot \| \mathbf{a}_2 \cdots \mathbf{a}_n \|$$

*has the $p$-invariant $C_p(U) = (-1, -1)_p$.*

LEMMA 1.6: *So long as we restrict ourselves to rational vectors, the $p$-invariant of a vector set, i.e., the $p$-invariant of the Gramian of the set is uniquely determined by the linear subspace generated by the vectors of the set.*

LEMMA 1.7: *For a matrix $A$ of the form*

$$A = eI_n + fG_n,$$

*where $I_n$ is the unit matrix of order n and $G_n$ is the $n \times n$ matrix whose elements are all unity,*

$$(1.9)\quad C_p(A) = (-1, -1)_p(-1, e)_p^{\frac{n(n-1)}{2}} (-1, g)_p(n, g)_p(n, e)_p(g, e)_p^{n-1},$$

*where we have put*

$$(1.10)\qquad g = e + nf.$$

Next we shall summarize the necessary properties of Hilbert's symbol [4, 6] and some of the fundamental properties of the Legendre symbol $(a/p)$ of the quadratic residue [6].

First of all, from the definition of $(a, b)_p$, it is clear that

$$(1.11)\qquad (a, b)_p = (b, a)_p,$$

and for any rational numbers $t$ and $u$,

$$(1.12)\qquad (at^2, bu^2)_p = (a, b)_p.$$

Hence in any calculation handling the Hilbert symbol, the square part of any rational number can be replaced by 1.

$$(1.13)\qquad \begin{aligned}(a, -a)_p &= +1\\ (a, a)_p &= (-1, a)_p\\ (a, b_1b_2)_p &= (a, b_1)_p(a, b_2)_p\end{aligned}$$

and [2, 4]

$$(1.14)\qquad (a, b)_p = (-ab, a + b)_p.$$

As a special case of (1.14), we have for every positive integer $n$:

$$(1.15)\qquad (n, n + 1)_p = (-1, n + 1)_p.$$

$$(1.16)\qquad \text{If}\quad (ab, p) = 1,\quad \text{then}\quad (a, b)_p = +1.$$

$$(1.17)\qquad \text{For an odd prime } p,\quad (p, a)_p = (a/p).$$

For the even prime 2, we have

$$(1.18)\qquad \begin{aligned}(p, q)_2 = (-1)^{\frac{p-1}{2}\frac{q-1}{2}},\qquad (2, p)_2 = (2/p),\qquad (-1, p)_2 = (-1/p),\\ (-1, 2)_2 = +1,\qquad (-1, -1)_2 = -1.\end{aligned}$$

And for $p = \infty$, we have

$$(1.19)\qquad \begin{aligned}(p, q)_\infty = (-1, 1)_\infty &= (2, p)_\infty\\ &= (-1, p)_\infty = +1,\qquad (-1, -1)_\infty = -1.\end{aligned}$$

In the above and hereafter, $p$ and $q$ denote odd primes.

For the Legendre symbol, the following properties are fundamental:

$$(1.20) \qquad (a/p) = (b/p) \quad \text{if} \quad a \equiv b (\mathrm{mod}\, p),$$

$$(1.21) \qquad (ab/p) = (a/p)(b/p)$$

and the reciprocity law [6]

$$(1.22) \qquad \left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}$$

supplemented by

$$(1.23) \qquad \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}, \qquad \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

298

**2. A P.B.I.B. design of triangular type.** Triangular association is defined as follows: The number of elements is $v = n(n-1)/2$, where $n$ is a positive integer. We take an $n \times n$ square, and fill the $n(n-1)/2$ positions above the main diagonal by the different elements, taken in order. The positions in the main diagonal are left blank, while the positions below the main diagonal are filled so that the scheme is symmetrical with respect to the main diagonal. Two elements in the same column are 1st associates, whereas two elements which do not occur in the same column are 2nd associates.

In this association each element has $n_i$ $i$th associates, where

$$n_1 = 2n - 4, \qquad n_2 = \frac{(n-2)(n-3)}{2}.$$

The parameters of association are as follows:

$$p_{11}^1 = n - 2, \qquad p_{12}^1 = n - 3 = p_{21}^1, \qquad p_{22}^1 = \frac{(n-3)(n-4)}{2},$$

$$p_{11}^2 = 4, \qquad p_{12}^2 = n - 8 = p_{21}^2, \qquad p_{22}^2 = \frac{(n-4)(n-5)}{2}.$$

Let the association matrices be $A_0 = I_v$, $A_1$, $A_2$, then it is known that these matrices generate a commutative linear associative algebra $\varkappa$ of rank 3, and the regular representation $(\varkappa)$ is given [7] by

$$(2.1) \quad (\varkappa): \begin{cases} A_0 \to I_3, \\ A_1 \to \mathcal{P}_1 = \begin{Vmatrix} 0 & 1 & 0 \\ 2n-4 & n-2 & 4 \\ 0 & n-3 & 2n-8 \end{Vmatrix}, \\ A_2 \to \mathcal{P}_2 = \begin{Vmatrix} 0 & 0 & 1 \\ 0 & n-3 & (n-8) \\ \dfrac{(n-2)(n-3)}{2} & \dfrac{(n-3)(n-4)}{2} & \dfrac{(n-4)(n-5)}{2} \end{Vmatrix}, \end{cases}$$

This regular representation $(\kappa)$ decomposes into three non-equivalent linear representations

$$(2.2)\quad \begin{array}{lll} \kappa_0 : A_0 \to 1, & A_1 \to n_1 = 2n - 4, & A_2 \to n_2 = (n-2)(n-3)/2, \\ \kappa_1 : A_0 \to 1, & A_1 \to n - 4, & A_2 \to -(n-3), \\ \kappa_2 : A_0 \to 1, & A_1 \to -2, & A_2 \to 1, \end{array}$$

having respective multiplicities

$$(2.3)\quad \alpha_0 = 1, \quad \alpha_1 = n - 1, \quad \alpha_2 = n(n-3)/2$$

in the algebra $\kappa$.

Suppose that we are given $v$ treatments having triangular association among them and $b$ blocks each having $k$ experimental units in such a way that

(1) each block contains $k$ different treatments,

(2) each treatment occurs in $r$ blocks, and

(3) any two treatments occur together in $\lambda_i$ blocks, if they are $i$th associates.

This design is called a P.B.I.B. *design of triangular type.*

If the incidence matrix of this design is denoted by $N$, it is also well known ([7], [8]) that

$$(2.4)\quad NN' = rA_0 + \lambda_1 A_1 + \lambda_2 A_2 .$$

Hence $NN'$ has eigenvalues

$$(2.5)\quad \begin{aligned} \rho_0 &= r + (2n-4)\lambda_1 + \frac{(n-2)(n-3)}{2}\lambda_2 = rk, \\ \rho_1 &= r + (n-4)\lambda_1 - (n-3)\lambda_2, \\ \rho_2 &= r - 2\lambda_1 + \lambda_2, \end{aligned}$$

with multiplicities 1, $(n - 1)$ and $n(n - 3)/2$ respectively.

It can be shown from the elements of linear associative algebra [9] that there exist three mutually orthogonal and symmetric idempotents $A_0^* = (1/v)G_v$, $A_1^*$, and $A_2^*$ with respective ranks 1, $n - 1$, and $n(n - 3)/2$, such that

$$(2.6)\quad NN' = \rho_0 A_0^* + \rho_1 A_1^* + \rho_2 A_2^*.$$

The column vectors of $A_i^*$ generate the eigenspace of $NN'$ corresponding to the eigenvalue $\rho_i$. Let us assume, without any loss of generality, that

$$\mathbf{a}_1^{0*}, \mathbf{a}_2^{1*}, \cdots, \mathbf{a}_n^{1*}, \mathbf{a}_{n+1}^{2*}, \cdots, \mathbf{a}_v^{2*}$$

are linearly independent, and let us put

$$(2.7)\quad S = \| \mathbf{a}_1^{0*} \mathbf{a}_2^{1*} \cdots \mathbf{a}_n^{1*} \mathbf{a}_{n+1}^{2*} \cdots \mathbf{a}_v^{2*} \|,$$

then $S$ is a non-singular $v \times v$ matrix with rational elements. Further let

$$(2.8)\quad Q_1 = \left\| \begin{matrix} \mathbf{a}_2^{1*\prime} \\ \vdots \\ \mathbf{a}_n^{1*\prime} \end{matrix} \right\| \, \| \mathbf{a}_2^{1*} \cdots \mathbf{a}_n^{1*} \|, \quad \text{and} \quad Q_2 = \left\| \begin{matrix} \mathbf{a}_{n+1}^{2*\prime} \\ \vdots \\ \mathbf{a}_v^{2*\prime} \end{matrix} \right\| \, \| \mathbf{a}_{n+1}^{2*} \cdots \mathbf{a}_v^{2*} \|,$$

then from (2.6) it follows that

$$S'NN'S = \left\| \begin{matrix} \rho_0 a_{11}^{0*} & 0 & 0 \\ 0 & \rho_1 Q_1 & 0 \\ 0 & 0 & \rho_2 Q_2 \end{matrix} \right\|,$$

or

$$NN' \sim \left\| \begin{matrix} \frac{rk}{v} & 0 & 0 \\ 0 & \rho_1 Q_1 & 0 \\ 0 & 0 & \rho_2 Q_2 \end{matrix} \right\|. \tag{2.9}$$

Since

$$S'S = \left\| \begin{matrix} \frac{1}{v} & 0 & 0 \\ 0 & Q_1 & 0 \\ 0 & 0 & Q_2 \end{matrix} \right\|,$$

we get

$$v \mid Q_1 \mid\mid Q_2 \mid \sim 1. \tag{2.10}$$

It has been shown by Corsten [10] that

$$\left\| \begin{matrix} \frac{1}{v} & 0 \\ 0 & Q_1 \end{matrix} \right\| \sim \left\| \begin{matrix} n-1 & 1 & \cdots & 1 \\ 1 & n-1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & n-1 \end{matrix} \right\|, \tag{2.11}$$

hence

$$\mid Q_1 \mid \sim n(n-2)^{n-1}. \tag{2.12}$$

**3. Necessary conditions for the existence of a regular symmetrical** P.B.I.B. **design of triangular type.** In this section, we shall show the non-existence of certain regular symmetrical P.B.I.B. designs of triangular type.

If the design is symmetrical, i.e., $v = b$ and $r = k$, then the incidence matrix $N$ is a square matrix with elements 0 and 1, hence in the regular case $\mid NN' \mid$ must be a perfect square. Thus first of all

$$\rho_1^{n-1}\rho_2^{\frac{1}{2}n(n-3)} = [r + (n-4)\lambda_1 - (n-3)\lambda_2]^{n-1}[r - 2\lambda_1 + \lambda_2]^{\frac{1}{2}n(n-3)} \sim 1 \tag{3.1}$$

and then, since $NN' \sim I_v$, we have

$$C_p(NN') = (-1, -1)_p \tag{3.2}$$

for all primes $p$. (3.1) and (3.2) are necessary conditions for the existence.

Now, from (2.9) we get

$$\begin{aligned} C_p(NN') = (-1, -1)_p(-1, v)_p(v, \rho_1^{n-1}\rho_2^{\frac{1}{2}n(n-3)} \mid Q_1 \mid\mid Q_2 \mid)_p \\ (\rho_1^{n-1} \mid Q_1 \mid, \rho_2^{\frac{1}{2}n(n-3)} \mid Q_2 \mid)_p\, C_p(\rho_1 Q_1)\cdot C_p(\rho_2 Q_2). \end{aligned} \tag{3.3}$$

By Lemma 1.4,

$$C_p(\rho_1 Q_1) = (-1, \rho_1)_p^{\frac{n(n-1)}{2}} (\rho_1, |Q_1|)_p^{n-2} C_p(Q_1) \tag{3.4}$$

$$C_p(\rho_2 Q_2) = (-1, \rho_2)_p^{\frac{n(n-1)(n-2)(n-3)}{8}} (\rho_2, |Q_2|)_p^{\frac{1}{2}n(n-3)-1} C_p(Q_2). \tag{3.5}$$

Since

$$v\,|Q_1|\,|Q_2| \sim 1, \qquad \rho_1^{n-1}\rho_2^{\frac{1}{2}n(n-3)} \sim 1,$$

it follows that

$$(v, \rho_1^{n-1}\rho_2^{\frac{1}{2}n(n-3)}\,|Q_1|\,|Q_2|)_p = (v, v)_p = (-1, v)_p, \tag{3.6}$$

$$\begin{aligned}(\rho_1^{n-1}\,|Q_1|, \rho_2^{\frac{1}{2}n(n-3)}\,|Q_2|)_p &= (\rho_1^{n-1}|Q_1|, \rho_1^{n-1}|Q_2|)_p \\ &= (\rho_1, v)_p^{n-1}(-1, \rho_1)_p^{n-1}(|Q_1|, |Q_2|)_p\end{aligned} \tag{3.7}$$

and

$$\begin{aligned}(\rho_2, |Q_2|)_p^{\frac{1}{2}n(n-3)-1} &= (\rho_2, |Q_2|)_p(\rho_1, v\,|Q_1|)_p^{n-1} \\ &= (\rho_2, |Q_2|)_p(\rho_1, v)_p^{n-1}(\rho_1, |Q_1|)_p^{n-1}.\end{aligned} \tag{3.8}$$

Substituting (3.4) to (3.8) into (3.3), we get

$$\begin{aligned}C_p(NN') = {} & (-1, -1)_p(\rho_1, v)_p^{n-1}(-1, \rho_1)_p^{n-1}(|Q_1|, |Q_2|)_p \\ & \cdot(-1, \rho_1)_p^{\frac{n(n-1)}{2}}(\rho_1, |Q_1|)_p^{n-2}(-1, \rho_2)_p^{\frac{n(n-1)(n-2)(n-3)}{8}} \\ & \cdot(\rho_2, |Q_2|)_p(\rho_1, v)_p^{n-1}(\rho_1, |Q_1|)_p^{n-1}C_p(Q_1)C_p(Q_2) \\ = {} & (-1, -1)_p(-1, \rho_1)_p^{\frac{(n-1)(n-2)}{2}}(-1, \rho_2)^{\frac{n(n-1)(n-2)(n-3)}{8}}(\rho_1, |Q_1|)_p \\ & \cdot(\rho_2, |Q_2|)_p(|Q_1|, |Q_2|)_p C_p(Q_1)\,C_p(Q_2),\end{aligned}$$

whereas by Lemma 1.5

$$(|Q_1|, |Q_2|)_p C_p(Q_1) C_p(Q_2) = +1$$

and

$$|Q_1| \sim n(n-2)^{n-1}, \qquad |Q_2| \sim 2(n-1)(n-2)^{n-1},$$

therefore

$$\begin{aligned}C_p(NN') = {} & (-1, -1)_p(-1, \rho_1)_p^{\frac{(n-1)(n-2)}{2}}(\rho_1, n)_p(\rho_1, n-2)_p^{n-1} \\ & \cdot(-1, \rho_2)_p^{\frac{n(n-1)(n-2)(n-3)}{8}}(\rho_2, 2)_p(\rho_2, n-1)_p(\rho_2, n-2)_p^{n-1}.\end{aligned} \tag{3.9}$$

Consequently (3.2) becomes

$$\begin{aligned}O_p \equiv {} & (-1, \rho_1)_p^{\frac{(n-1)(n-2)}{2}}(\rho_1, n)_p(\rho_1, n-2)_p^{n-1}(-1, \rho_2)_p^{\frac{n(n-1)(n-2)(n-3)}{8}} \\ & \cdot(\rho_2, 2)_p(\rho_2, n-1)_p(\rho_2, n-2)_p^{n-1} = +1\end{aligned} \tag{3.10}$$

for all primes $p$.

**4. Examples of non-existent P.B.I.B. designs of triangular type.**

(1) $n = 7; \quad v = b = 21, \quad r = k = 6. \quad \lambda_1 = 0, \quad \lambda_2 = 3$

$$\rho_1 = -6, \quad \rho_2 = 9$$

$$O_p = (-1, 6)_p(-6, 7)_p = (-1, -1)_p(-1, 2)_p(-1, 3)_p(-1, 7)_p(2, 7)_p(3, 7)_p$$

$$O_3 = \left(\frac{-1}{3}\right)\left(\frac{7}{3}\right) = -1.$$

Hence this design is impossible.

(2) $n = 7; \quad v = b = 21, \quad r = k = 10, \quad \lambda_1 = 0, \quad \lambda_2 = 9$

$$\rho_1 = -26, \quad \rho_2 = 19$$

$$O_p = (-1, -26)_p(-26, 7)_p(19, 2)_p(19, 6)_p(-1, 19)_p$$
$$= (-1, -1)_p(-1, 2)_p(-1, 13)_p(-1, 7)_p(2, 7)_p(13, 7)_p(19, 3)_p(-1, 19)_p$$

$$O_{13} = \left(\frac{-1}{13}\right)\left(\frac{7}{13}\right) = \left(\frac{13}{7}\right) = \left(\frac{-1}{7}\right) = -1.$$

Hence this design is impossible.

(3) $n = 7; \quad v = b = 21, \quad r = k = 10, \quad \lambda_1 = 1, \quad \lambda_2 = 8$

$$\rho_1 = -19, \quad \rho_2 = 16$$

$$O_p = (-1, -19)_p(-19, 7)_p = (-1, -1)_p(-1, 19)_p(-1, 7)_p(19, 7)_p$$

$$O_{19} = \left(\frac{-1}{19}\right)\left(\frac{7}{19}\right) = \left(\frac{19}{7}\right) = \left(\frac{5}{7}\right) = \left(\frac{2}{5}\right) = -1.$$

Hence this design is impossible.

(4) $n = 7; \quad v = b = 21, \quad r = k = 10, \quad \lambda_1 = 2, \quad \lambda_2 = 7$

$$\rho_1 = -12, \quad \rho_2 = 13$$

$$O_p = (-1, -12)_p(-12, 7)_p(13, 2)_p(13, 6)_p(-1, 13)_p$$
$$= (-1, -1)_p(-1, 3)_p(-1, 7)_p(3, 7)_p(13, 3)_p(-1, 13)_p$$

$$O_3 = \left(\frac{-1}{3}\right)\left(\frac{7}{3}\right)\left(\frac{13}{3}\right) = \left(\frac{-1}{3}\right) = -1.$$

Hence this design is impossible.

(5) $n = 7; \quad v = b = 21, \quad r = k = 10, \quad \lambda_1 = 3, \quad \lambda_2 = 6$

$$\rho_1 = -5, \quad \rho_2 = 10$$

$$O_p = (-1, -5)_p(-5, 7)_p(10, 2)_p(10, 6)_p(-1, 10)_p$$
$$= (-1, -1)_p(5, 7)_p(-1, 7)_p(2, 7)_p(2, 3)_p(5, 3)_p$$

$$O_2 = (-1, -1)_2(5, 7)_2(-1, 7)_2(2, 7)_2(2, 3)_2(5, 3)_2 = -1.$$

Hence this design is impossible.

(6) $n = 7;\quad v = b = 21,\quad r = k = 10,\quad \lambda_1 = 8,\quad \lambda_2 = 1$

$$\rho_1 = 30, \quad \rho_2 = -5$$

$$O_p = (-1, 30)_p(7, 30)_p(-1, -5)_p(2, -5)_p(-5, 6)_p$$

$$O_3 = \left(\frac{-1}{3}\right)\left(\frac{7}{3}\right)\left(\frac{-5}{3}\right) = \left(\frac{-1}{3}\right) = -1.$$

Hence this design is impossible.

## REFERENCES


[1] S. S. Shrikhande, "The impossibility of certain symmetrical balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 106–111.

[2] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[3] S. S. Shrikhande, "The uniqueness of the $L_2$ association scheme," *Ann. Math. Stat.*, Vol. 30 (1959) pp. 781–798.

[4] B. W. Jones, *The Arithmetic Theory of Quadratic Forms*, The Mathematical Association of America, John Wiley and Sons, Inc., 1950.

[5] H. Hasse, "Über die Äquivalenz quadratischer Formen im Körper der rationalen Zahlen," *J. Reine Angew. Math.*, Vol. 152 (1923), pp. 205–224.

[6] K. Hensel, *Zahlentheorie*, G. J. Göschen'sche Verlagshandlung G.m.b.H., Berlin and Leipzig, 1915.

[7] R. C. Bose and D. M. Mesner, "On linear associative algebras corresponding to association schemes of partially balanced designs," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 21–38.

[8] W. S. Connor and W. H. Clatworthy, "Some theorems for partially balanced designs," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–112.

[9] J. Ogawa, "Linear associative algebra and some of its applications to the theory of design of experiments," unpublished lecture notes at University of North Carolina, 1958.

[10] L. C. A. Corsten, "Proper spaces related to triangular P.B.I.B. designs", to appear in *Ann. Math. Stat.*

# OPTIMAL SPACING IN REGRESSION ANALYSIS[1]

By H. A. David and Beverly E. Arens

*Virginia Polytechnic Institute*

**1. Introduction and summary.** When a response (or dependent) variable $y$ can be observed for a continuous range of values of the independent variable $x$, which is at the control of the experimenter, the question arises as to how a given number of observations should be spaced. It will be assumed that $x$ is measurable without error and that $y$ differs from the true response function $f(x)$ by a random term $z$ with mean zero and constant variance $\sigma^2$. We suppose that the aim of the experimenter is to estimate $f(x)$, or possibly the mean response $\overline{f(x)}$, on the basis of $n$ observations $(x_i, y_i)$.

Various aspects of this problem of optimal spacing have been studied for the case where $f(x)$ is known apart from some parameters (see e.g. Elfving [3], Chernoff [1], de la Garza [2], and Kiefer and Wolfowitz [8]). However, the functional form of $f(x)$ is often unknown or only approximately known. In the absence of a specific model to the contrary, polynomial approximations to $f(x)$ provide a convenient approach. Section 2 deals briefly with the non-statistical case $\sigma = 0$ when the problem of choosing $n$ abscissae in order to approximate to $f(x)$ by a polynomial of degree $n - 1$ reduces to one of optimum interpolation and that of integrating $f(x)$ reduces to Gaussian quadrature. For a fuller account of this part see Hildebrand [5] or Kopal [6].

If the response contains a random element, a polynomial of degree $n - 1$ or less may be fitted to the $n$ observations by least squares. The error of approximation will now be due, in general, both to random error and the use of an incorrect approximating function. We confine ourselves to the case of fitting a straight line when the true response, while roughly linear, may contain a quadratic component. Two criteria are considered in arriving at the two abscissae resulting in an optimal fit. The first of these criteria ((3.2) below) has also been discussed in a recent paper by Box and Draper [7] who have extended its use to the case of several independent variables.

It is shown in Section 6 that for $x$-values symmetrically spaced about the centre of the region of interest nothing is gained in fitting a straight line by the use of more than two such abscissae. These optimal abscissae are determined in Sections 3 and 4.

The emphasis of the present approach is on attaining an optimal straight line fit with a small number of observations, rather than on detecting departures from linearity. For the latter purpose more than two abscissae would, of course, be needed, but the number of observations required may well be uneconomically large. In Section 7 comparisons with some other simple spacings are made.

---

Received August 25, 1958; revised February 18, 1959.

[1] This paper was prepared with the support of the Office of Ordnance Research, U.S. Army, under Contract DA-36-034-ORD-1527 RD.

As an illustration, consider the calibration of a large number of instruments for a range of $x$ in which $f(x)$ is known to be approximately linear. In this case adequate accuracy may be attainable by the use of two observations only. If $\sigma$ is not negligibly small several observations may be taken at each of two appropriately selected settings, especially if it is much easier to repeat measurements at a given setting than to turn to a new one (compare de la Garza [2]).

An example illustrating the methods proposed is given in Section 8.

**2. Optimal spacing in the absence of random error.** We suppose that the region of interest of the independent variable is finite and that it has been transformed into the closed interval $(-1, 1)$. If $g_{n-1}(x)$, a polynomial of degree $n - 1$, agrees with $f(x)$ at the $n$ abscissae $x_1, x_2, \cdots x_n$, and if $f(x)$ has $n$ continuous derivatives in $(-1, 1)$ the remainder $R(x) = f(x) - g_{n-1}(x)$ may be expressed as

$$R(x) = \pi(x) \frac{f^{(n)}(\xi)}{n!},$$

where $\pi(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$, and $|\xi| < 1$. In order to make $g_{n-1}(x)$ a desirable approximating function it is natural to attempt to minimize $|R(x)|$ in some sense by an appropriate choice of abscissae. However, $\xi$ depends, in general, not only on the abscissae but also on $x$ and the nature of the function $f(x)$. It is therefore customary to content oneself with the minimization, in the sense chosen, of $|\pi(x)|$. If $f(x)$ is a polynomial of degree $n$, $|R(x)|$ will also be minimized, but more generally the minimization of $|R(x)|$ will be only approximate (compare [5], Section 9.6).

We consider the following two alternative requirements:

$$\int_{-1}^{1} \pi^2(x)\, dx = \min, \tag{2.1}$$

$$\max_{(-1,1)} |\pi(x)| = \min. \tag{2.2}$$

The first is a criterion of closest overall fit and gives the abscissae as the $n$ zeros of the Legendre polynomial $P_n(x)$ of degree $n$; the second results in abscissae which are the zeros of the Tchebysheff polynomial $T_n(x) = \cos(n \cos^{-1} x)$. Corresponding to these two cases we shall speak of Legendre and Tchebysheff spacing. Generally, the latter would be regarded as more appropriate in the problem of calibration outlined in the introduction.

Criteria (2.1) and (2.2) may also be given a statistical interpretation. To this end we note that (2.2) can be shown (e.g. [5], Section 9.6) to be equivalent to

$$\int_{-1}^{1} \frac{\pi^2(x)}{(1 - x^2)^{\frac{1}{2}}}\, dx = \min. \tag{2.3}$$

Suppose $g_{n-1}(x)$ is required for a value of $x$ chosen randomly in $(-1, 1)$. Then, clearly, $\mathcal{E}[\pi^2(x)]$ is minimized by (2.1) if $x$ is uniformly distributed in $(-1, 1)$ and by (2.2) if $\cos^{-1} x$ is uniformly distributed in $(0, \pi)$.

A further advantage of the above spacings is that the integral approximation

$$\int_{-1}^{1} w(x)f(x)\,dx \doteq \int_{-1}^{1} w(x)g_{n-1}(x)\,dx \tag{2.4}$$

is a Gaussian quadrature formula with weight function $w(x) = 1$ for Legendre spacing and $w(x) = (1 - x^2)^{-\frac{1}{2}}$ for Tchebysheff spacing (see e.g. [6], Chapter VII). Thus if the integral of $f(x)$ over $(-1, 1)$ is required it is given by

$$\sum_{k=1}^{n} H_k f(x_k) + E_n, \tag{2.5}$$

where the $H_k$ are tabulated weights, the $x_k$ are the zeros of the Legendre polynomial of degree $n$, and the error of integration $E_n$ is given by

$$\frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3} f^{(2n)}(\eta). \tag{2.6}$$

The integration formula (2.5), although it uses only $n$ ordinates, is therefore of degree of precision $2n - 1$, i.e., the integration is exact if $f(x)$ is a polynomial of degree $2n - 1$ or less. For a general function $f(x)$, (2.5) can be shown to be optimal in the sense that the coefficient of $f^{(2n)}(\eta)$ in (2.6) is smaller than for any other integration formula of degree of precision $2n - 1$.

**3. Criteria for optimal spacing in the presence of random error in the observed response.** We take the observed response to be

$$y(x) = f(x) + z$$

where $f(x)$ is the true response and $z$ is a variate with zero mean and variance $\sigma^2$ independent of $x$. As stated in the introduction we shall consider specifically the case where $f(x)$ is a quadratic while the fitted curve is a straight line. We suppose that $\frac{1}{2}n$ observations are taken at each of $x_1$, $x_2 (x_1 < x_2)$ and that the corresponding observed mean responses are $\bar{y}_1$, $\bar{y}_2$. The use of more than two abscissae is discussed in Section 6.

The fitted straight line is then

$$Y(x) = \hat{c}_0 + \hat{c}_1(x - \bar{x}),$$

where

$$\hat{c}_0 = \bar{y}, \qquad \hat{c}_1 = (\bar{y}_2 - \bar{y}_1)/(x_2 - x_1). \tag{3.1}$$

For $\sigma = 0$ we know from Section 2 that taking $x_1$, $x_2$ as the zeros of $P_2(x) = \frac{1}{2}(3x^2 - 1)$ or of $T_2(x) = 2x^2 - 1$ will minimize respectively

$$\int_{-1}^{1} [f(x) - Y(x)]^2\,dx,$$

$$\max_{(-1,1)} |f(x) - Y(x)|.$$

Of course, in this case we would take $n = 2$.

If $\sigma \neq 0$ it is a natural extension to try to choose $x_1$, $x_2$ so as to minimize

respectively the expected mean square error $\bar{E}$ given by

$$(3.2) \qquad \bar{E} = \frac{1}{2}\,\mathcal{E}\int_{-1}^{1}[f(x) - Y(x)]^2\,dx = \frac{1}{2}\int_{-1}^{1}\mathcal{E}[f(x) - Y(x)]^2\,dx$$

or the *maximum expected squared error*

$$(3.3) \qquad E_{\max} = \max_{(-1,1)} \mathcal{E}[f(x) - Y(x)]^2.$$

These criteria are equally applicable to the case where $f(x)$ is a polynomial of degree $p \leqq n$ while $Y(x)$ is of degree $p - 1$, there being $n$ locations. If $f(x)$ and $Y(x)$ are of the same degree, (3.3) reduces to the minimization of the maximum variance which has been considered by de la Garza [2], Guest [4], and Kiefer and Wolfowitz [8].

**4. Legendre and Tchebysheff spacing for $\sigma \neq 0$.** Before obtaining the abscissae $x_1$, $x_2$ satisfying (3.2) or (3.3) we consider briefly the effects of using Legendre or Tchebysheff spacing when $\sigma \neq 0$. For the former case it is convenient to express $f(x)$ in terms of Legendre polynomials, viz.,

$$f(x) = c_0 + c_1P_1(x) + c_2P_2(x).$$

Then for any two symmetrical locations $(-x_1 = x_2)$ we have from (3.1)

$$(4.1) \qquad \mathcal{E}(\hat{c}_0) = c_0 + c_2P_2(x_2), \qquad \mathcal{E}(\hat{c}_1) = c_1$$

and

$$(4.2) \qquad \operatorname{var}\hat{c}_0 = \frac{\sigma^2}{n}, \qquad \operatorname{var}\hat{c}_1 = \frac{\sigma^2}{nx_2^2}, \qquad \operatorname{cov}(\hat{c}_0, \hat{c}_1) = 0.$$

Thus, if $x_2 = 1/\sqrt{3}$, $\hat{c}_0$ and $\hat{c}_1$ are unbiased estimators of $c_0$ and $c_1$. In this case

$$\mathcal{E}[f(x) - Y(x)] = c_2P_2(x)$$

and

$$(4.3) \qquad \int_{-1}^{1}\mathcal{E}[f(x) - Y(x)]\,dx = 0.$$

Thus $Y(x)$ may be said to be "unbiased on the average" as an estimator of $f(x)$. Interchanging the integration and expectation signs in (4.3) we see that the expected area under $Y(x)$ is equal to the area under $f(x)$, a result which continues to be true if $f(x)$ is a cubic, in line with the optimal integration properties of Legendre spacing. With Legendre spacing we have also

$$\begin{aligned}\mathcal{E}[f(x) - Y(x)]^2 &= \operatorname{var}\hat{c}_0 + x^2\operatorname{var}\hat{c}_1 + c_2^2P_2^2(x)\\ &= \tfrac{1}{2}\sigma'^2 + \tfrac{3}{2}x^2\sigma'^2 + c_2^2P_2^2(x),\end{aligned}$$

where $\sigma'^2 = 2\sigma^2/n$, so that the expected mean square error is

$$(4.4) \qquad \bar{E}_L = \sigma'^2 + \tfrac{1}{5}c_2^2.$$

The results (4.1), (4.2) but not (4.3) hold also with obvious changes when $f(x)$ is expressed in terms of Tchebysheff polynomials, viz.,

$$f(x) = b_0 + b_1T_1(x) + b_2T_2(x).$$

In this case $\hat{c}_0$ and $\hat{c}_1$ are unbiased estimators of $b_0$ and $b_1$ if $x_2 = 1/\sqrt{2}$.

**5. Optimal spacing with two locations.** We consider first the minimization of $\bar{E}$ in (3.2) and to this end show that the search for optimal values of $x_1$ and $x_2$ may be confined to the symmetrical spacing $-x_1 = x_2$.

In place of (4.1) and (4.2) we now have

$$\mathcal{E}(\hat{c}_0) = c_0 + c_1\bar{x} + c_2\bar{P}_2(x), \qquad \mathcal{E}(\hat{c}_1) = c_1 + 3c_2\bar{x}$$

and

$$\operatorname{var} \hat{c}_0 = \tfrac{1}{2}\sigma'^2, \qquad \operatorname{var} \hat{c}_1 = \frac{2\sigma'^2}{(x_2 - x_1)^2}, \qquad \operatorname{cov}(\hat{c}_0, \hat{c}_1) = 0,$$

where

$$\bar{P}_2(x) = \tfrac{1}{2}[P_2(x_1) + P_2(x_2)].$$

It follows that

$$\mathcal{E}[f(x) - Y(x)] = c_2[P_2(x) - \bar{P}_2(x) - 3\bar{x}(x - \bar{x})]$$

and

$$(5.1) \quad \mathcal{E}[f(x) - Y(x)]^2 = \tfrac{1}{2}\sigma'^2 + \frac{2\sigma'^2}{(x_2 - x_1)^2}(x - \bar{x})^2 + \{\mathcal{E}[f(x) - Y(x)]\}^2.$$

Hence

$$(5.2) \quad \bar{E} = \tfrac{1}{2}\sigma'^2 + \frac{\sigma'^2}{3(x_2 - x_1)^2}[(1 - \bar{x})^3 + (1 + \bar{x})^3] + c_2^2 X,$$

where

$$X = \{\tfrac{1}{5} + \bar{P}_2^2(x) + \tfrac{3}{2}\bar{x}^2[(1 - \bar{x})^3 + (1 + \bar{x})^3] - 6\bar{P}_2(x)\bar{x}^2\}.$$

Let $x_2 - x_1 = 2\alpha$; then $|\bar{x}| \leqq 1 - \alpha$. Writing also $\bar{x} = y, x_1 = y - \alpha, x_2 = y + \alpha$, we have

$$X = \tfrac{1}{5} + \tfrac{1}{4}(3y^2 + 3\alpha^2 - 1)^2 + 6y^2 - 9y^2\alpha^2,$$

and for any given $\alpha$ this may be shown to have a single minimum at $y = 0$ provided $|y| \leqq 1 - \alpha$, $|\alpha| < 1$. Corresponding to any given $\alpha$, therefore, $X$ and hence $\bar{E}$ are minimized by taking $x_1 = -\alpha$, $x_2 = \alpha$.

From (5.2) we may now write

$$(5.2') \quad \bar{E} = \tfrac{1}{2}\sigma'^2 + \frac{\sigma'^2}{6x_2^2} + c_2^2[\tfrac{1}{5} + P_2^2(x_2)].$$

This is to be minimized with respect to $x_2$. Setting $d\bar{E}/dx_2 = 0$ we find $x_2$ to be a root of the equation

TABLE 1

*Values of $-x_1 = x_2$, as a function of $b = \sigma'/|c_2|$, giving* (i) *generalized Legendre and* (ii) *generalized Tchebysheff spacing*

| $b$ | $-x_1 = x_2$ (i) | $-x_1 = x_2$ (ii) |
|---|---|---|
| 0 | 0.577 | 0.707 |
| 0.3 | .599 | .721 |
| 0.6 | .642 | .755 |
| 0.9 | .685 | .800 |
| 1.2 | .725 | .850 |
| 1.5 | .762 | .899 |
| 1.8 | .796 | .949 |
| 2.1 | .827 | .997 |
| 2.4 | .855 | 1.000 |
| 2.7 | .882 | ... |
| 3.0 | .908 | ... |
| 3.3 | .932 | ... |
| 3.6 | .955 | ... |
| 3.9 | .976 | ... |
| 4.2 | .997 | ... |
| 4.5 | 1.000 | 1.000 |

N.B. $x_2 = 1$ for $b \geqq 4.243$ in (i) and $b \geqq 2.121$ in (ii).

$$x_2^4(3x_2^2 - 1) = a, \tag{5.3}$$

where $a = \sigma'^2/(9c_2^2)$. Thus $x_2$ is a function of $a$ or equivalently, of $b = \sigma'/|c_2|$. Equation (5.3) is a cubic in $x_2^2$ with only one real root which corresponds to the required minimum. For $\sigma = 0$, (5.3) gives Legendre spacing. On the other hand, if $\sigma \neq 0$ but $c_2 = 0$, so that $a$ is infinite, $\bar{E}$ will be minimized by making $x_2$ as large as possible, i.e., $x_2 = 1$. In fact, $x_2^2 = 1$ for $a = 2$. For $a > 2$ or $\sigma'^2 > 18c_2^2$ we still take $x_2 = 1$. The dependence of $x_2$ on $b$ is shown in Table 1.

We turn now to the minimization of the maximum expected squared error of (3.3). In this case also we may take $-x_1 = x_2$. By (5.1) it is therefore required to maximize

$$X' = \frac{\sigma'^2 x^2}{2x_2^2} + \tfrac{9}{4}c_2^2(x^2 - x_2^2)^2$$

with respect to $x$ and subsequently to minimize this maximum with respect to $x_2$. If we regard $X'$ as a quadratic in $x^2$ for $0 \leqq x^2 \leqq 1$, it is clear that its maximum occurs at $x^2 = 0$ or 1. For $x^2 = 0$, $X'$ increases in $x_2$ from 0 to $(9/4)c_2^2$ while for $x^2 = 1$, $X'$ decreases from $\infty$ to $\frac{1}{2}\sigma'^2$. Thus if $\sigma'^2 \geqq (9/2)c_2^2$, then $x_2 = 1$ is the solution. Otherwise the solution is that value of $x_2$ between 0 and 1 which equalizes $X'$ for $x^2 = 0$ and $x^2 = 1$. This occurs for $x_2^4 - \frac{1}{2}x_2^2 = a$, so that for optimal spacing

(5.4) $$x_2 = \tfrac{1}{2}[1 + (1 + 16a)^{\frac{1}{2}}]^{\frac{1}{2}} \text{ or } 1,$$

whichever is smaller. For $\sigma = 0$, (5.4) gives Tchebysheff spacing The dependence of $x_2$ on $b$ in this case is also shown in Table 1.

The two types of spacing. may conveniently be referred to as generalized Legendre and generalized Tchebysheff spacing.

**6. Possible use of more than two locations.** Suppose that more than two locations are available to us and that we fit a least squares straight line to the $n$ observations. If these are taken at $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, it seems natural to continue to assume symmetry of spacing, i.e., $x_i = -x_{n+1-i}$ $(i = 1, 2, \cdots, n)$, so that both $\sum x_i$ and $\sum x_i^3$ vanish. Then

(6.1) $$\hat{c}_0 = \bar{y}, \qquad \hat{c}_1 = \sum x_i y_i / \sum x_i^2,$$

(6.2) $$\mathcal{E}(\hat{c}_0) = c_0 + \frac{c_2 \sum P_2(x_i)}{n}, \qquad \mathcal{E}(\hat{c}_1) = c_1,$$

(6.3) $$\operatorname{var} \hat{c}_0 = \frac{\sigma^2}{n}, \qquad \operatorname{var} \hat{c}_1 = \frac{\sigma^2}{\sum x_i^2}, \qquad \operatorname{cov}(\hat{c}_0, \hat{c}_1) = 0.$$

Now comparison of (6.2), (6.3) with (4.1), (4.2) shows that the two sets of equations become identical if $\sum x_i^2 = n x_2^2$. In other words, corresponding to any symmetrical configuration of $n$ locations, a value $x_2 (> 0)$ can be found such that $\frac{1}{2}n$ observations at each of $\pm x_2$ give estimators $\hat{c}_0$, $\hat{c}_1$ with the same expectations, variances and covariances. It follows that the two spacings are equivalent from this point of view as well as on the basis of any criteria depending on the first two moments only, such as (3.2) and (3.3). See also Box and Draper [7], who obtain similar results on merely taking $\sum x_i = 0$.

In certain situations it is advantageous to vary the independent variable as little as necessary. Apart from its convenience the use of two locations will obviously be optimal on this score also. Of course, more than two locations are necessary to detect departures from linearity in $f(x)$ but this is not our aim here.

As before, we have taken $n$ even which would be the usual situation. However, if $n$ is odd, the number of locations is reducible to three, and an odd number of observations has to be taken at $x = 0$. Clearly, the narrowest spacing is given by a single observation at $x = 0$ and $\frac{1}{2}(n - 1)$ observations at each of $\pm x_2[n/(n - 1)]^{\frac{1}{2}}$.

These equivalence results may be compared with those obtained by Elfving [3] and de la Garza [2] in the case when the fitted function and the true response are polynomials of the *same* degree, so that no bias enters. For $c_2 = 0$ the present result is a special case of theirs; the equivalence continues to hold for $c_2 \neq 0$ because by (6.2), (6.3) the bias in $\hat{c}_0$ is, like the variance of $\hat{c}_1$, a function of $\sum x_i^2$.

TABLE 2

*$\bar{E}$ and $E_{max}$ as functions of $b = \sigma'$, for various spacings*

$\bar{E}$

| $b$ | (i) Generalized Legendre | (ii) Legendre $(-1/\sqrt{3}, 1/\sqrt{3})$ | (iii) $(-1, 1)$ | (iv) $(-1, 0, 1)$ | (v) $(\pm 0.2, \pm 0.6, \pm 1)$ |
|---|---|---|---|---|---|
| 0 | 0.20 | 0.20 | 1.20 | 0.45 | 0.24 |
| 0.6 | 0.54 | 0.56 | 1.44 | 0.72 | 0.55 |
| 1.2 | 1.46 | 1.64 | 2.16 | 1.53 | 1.47 |
| 1.8 | 2.88 | 3.44 | 3.36 | 2.88 | 3.02 |
| 2.4 | 4.75 | 5.96 | 5.04 | 4.77 | 5.18 |
| 3.0 | 7.06 | 9.20 | 7.20 | 7.20 | 7.95 |
| 3.6 | 9.80 | 13.16 | 9.84 | 10.17 | 11.35 |
| 4.2 | 12.96 | 17.84 | 12.96 | 13.68 | 15.36 |
| 4.8 | 16.56 | 23.24 | 16.56 | 17.73 | 19.99 |

$E_{max}$

| $b$ | (i) Generalized Tchebysheff | (ii) Tchebysheff $(-1/\sqrt{2}, 1/\sqrt{2})$ | (iii) $(-1, 1)$ | (iv) $(-1, 0, 1)$ | (v) $(\pm 0.2, \pm 0.6, \pm 1)$ |
|---|---|---|---|---|---|
| 0 | 0.56 | 0.56 | 2.25 | 1.00 | 0.64 |
| 0.6 | 0.91 | 1.10 | 2.43 | 1.18 | 1.21 |
| 1.2 | 1.89 | 2.72 | 2.97 | 2.05 | 2.90 |
| 1.8 | 3.44 | 5.42 | 3.87 | 4.30 | 5.73 |
| 2.4 | 5.76 | 9.20 | 5.76 | 7.45 | 9.69 |

**7. Comparison of $\bar{E}$ and $E_{max}$ for various spacings.** It is of interest to compare our two optimal spacings with other simple spacings. The results of a number of such comparisons are set out in Table 2. For definiteness, and without real loss of generality, we have taken the true quadratic response as $f(x) = c_0 + c_1P_1(x) + P_2(x)$, so that $b = \sigma'$. For various values of $\sigma'$ Table 2 lists $\bar{E}$ which from (5.2′) and Section 6 is given by

$$\bar{E} = \sigma'^2(0.5 + 6\gamma^{-1}) + (0.45 - 1.5\gamma + 2.25\gamma^2),$$

where $\gamma = \sum x_i^2/n$; and also $E_{max}$ which is the larger of $0.5\sigma'^2 + 2.25\gamma^2$ and $0.5\sigma'^2(1 + \gamma^{-1}) + 2.25(1 - \gamma)^2$.

**8. An example.** To illustrate Legendre and generalized Legendre spacing we suppose that the true law under study is

$$h(x') = 8 - x' + \tfrac{1}{10}x'^2, \qquad 0 \leqq x' \leqq 10. \tag{8.1}$$

Put $x' = 5 + 5x$ to transform this to

$$f(x) = \tfrac{11}{2} - \tfrac{5}{2}x + \tfrac{5}{2}x^2, \qquad -1 \leqq x \leqq 1$$

$$= \tfrac{14}{3} - \tfrac{5}{2}P_1(x) + \tfrac{5}{3}P_2(x).$$

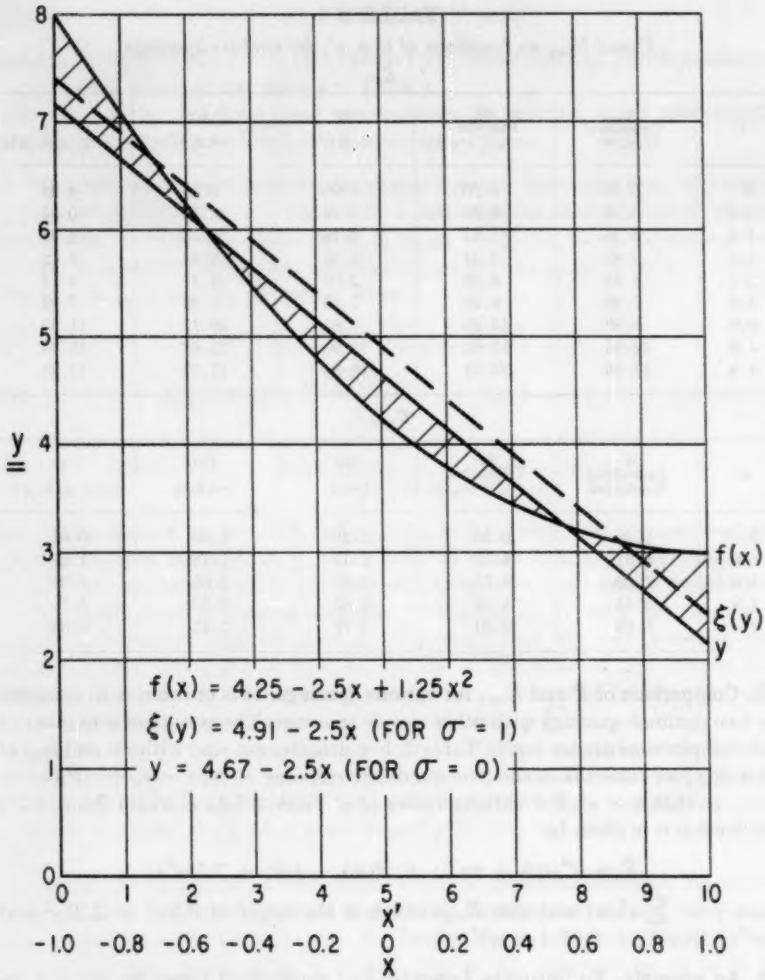


FIG. 1. Illustrating Legendre and generalized Legendre spacing

If this function may be observed for only two values of $x$ the closest overall fit in the case of no random error ($\sigma = 0$) is obtained by taking $-x_1 = x_2 = 1/\sqrt{3}$, which results in the straight line of approximation

$$Y(x) = \tfrac{14}{3} - \tfrac{5}{2}x.$$

The average error is zero and the mean square error is by (4.4) simply $c_2^2/5 = 5/36$.

When $\sigma \neq 0$ the optimal spacing is given by Table 1 with $b = 6\sigma'/5$. Thus for $\sigma' < \frac{1}{2}$ the spacing is only slightly wider than the Legendre spacing while for $\sigma' > 3$ the observations should be taken at $x = -1, 1$. The expected line still has slope $c_1 = -5/2$ but is displaced upwards through a vertical distance $c_2P_2(x_2)$.

For $\sigma' = 1$ the situation is shown in Figure 1. In this case $x_2 = 0.725$ and the expected mean square error $\bar{E} = 1.014$ by (5.2′) or from Table 2 ($1.46 \times (5/6)^2$). This may be compared with $\bar{E}_L = 1.139$. For $\sigma' = 2$, we have $x_2 = 0.855$ and $\bar{E} = 3.298$, $\bar{E}_L = 4.139$.

In this example we have taken $h(x')$ as known so that the results could be presented graphically. However, it is clear that the optimal locations are determined completely by the coefficient of $x'^2$ in (8.1), the specified range of $x'$ and the standard deviation $\sigma'$. Thus the same results hold approximately when all that is known is that the response function is linear in the range (0, 10) apart from a quadratic term with coefficient of the order 0.05.

**9. Acknowledgments.** We are grateful to James L. Dolby, who drew our attention to the kind of problem considered, to the authors of papers [7] and [8] for letting us see these in advance of publication, and to the referees for a number of very helpful suggestions.

## REFERENCES

[1] Herman Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 586–602.

[2] A. de la Garza, "Spacing of information in polynomial regression," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 123–130.

[3] G. Elfving, "Optimum allocation in linear regression theory," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–262.

[4] P. G. Guest, "The spacing of observations in polynomial regression," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 294–299.

[5] F. B. Hildebrand, *Introduction to Numerical Analysis*, McGraw-Hill, New York, 1956.

[6] Zdeněk Kopal, *Numerical Analysis*, John Wiley and Sons, New York, 1955.

[7] G. E. P. Box and Norman R. Draper, "A basis for the selection of a response surface design," Paper read at 31st Session, Int. Stat. Inst., September 1958.

[8] J. Kiefer and J. Wolfowitz, "Optimum designs in regression problems," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 271–294.

# THIRD ORDER ROTATABLE DESIGNS FOR EXPLORING RESPONSE SURFACES[1, 2]

By D. A. Gardiner

*Oak Ridge National Laboratory*

and

A. H. E. Grandage and R. J. Hader

*North Carolina State College*

**1. Introduction.** This paper considers a problem arising in the design of experiments for empirically investigating the relationship between a dependent and several independent variables, all variables being continuous. It is assumed that the form of the functional relationship is unknown but that within the range of interest, the function may be represented by a Taylor series expansion of moderately low order. Specifically, the problem considered herein is that choice of combinations of levels of the independent variables which, a) will enable an experimenter to approximate a functional relationship by fitting a Taylor series expansion through terms of order 3, by the method of least squares, and b) will have the property of rotatability. Such a choice of combinations of levels of the independent variables will be called a third order rotatable design.

For the sake of brevity, the abbreviation *d*th ORD will be used to denote *d*th order rotatable design.

**2. Rotatability.** The property of rotatability as a desirable quality of an experimental design was first advanced by Box and Hunter in [1]. This property is that the variances of estimates of the response made from the least squares estimates of the Taylor series are constant on circles, spheres or hyper-spheres about the center of the design. Thus, a rotatable design, that is, a design which achieves this property, could be rotated through any angle around its center and the variances of responses estimated from it would be unchanged.

Box and Hunter proved that a necessary and sufficient condition for a design of order $d$ ($d = 1, 2, 3, \cdots$) to be rotatable is that the moments of the independent variables be the same, through order $2d$, as those of a spherical distribution, or that these moments be invariant under a rotation of the design around its center.

Let $k$ be the number of independent variables, or factors, and let $x_{1u}, x_{2u}, \cdots, x_{ku}$ be the levels of these variables for the $u$th experimental point in the factor space, ($u = 1, 2, \cdots, N$). Then a $p$th order moment is defined as

$$N^{-1} \sum_{u=1}^{N} x_{1u}^{q} x_{2u}^{r} \cdots x_{ku}^{t},$$

Received June 16, 1958; revised March 16, 1959.

[1] This work was sponsored in part by the U. S. Army under Contract No. DA-36-034-ORD-1517(RD) (Experimental Designs for Industrial Research).

[2] This paper is based on a part of the Ph.D. Thesis (N. C. State College, 1956) by Gardiner [2].

where $0 \leqq q$, $0 \leqq r, \cdots, 0 \leqq t$, and $q + r \cdots + t = p$. Further, let the independent variables be standardized so that

$$\sum_{u=1}^{N} x_{iu}^2 = N \qquad (i = 1, 2, \cdots, k). \tag{2.1}$$

Let $\eta_u$ be the expectation of the response at the $u$th experimental point. For a polynomial equation of third order this may be written

$$\eta_u = \beta_0 + \sum_{i=1}^{k} \beta_i x_{iu} + \sum_{i \leqq j=1}^{k} \beta_{ij} x_{iu} x_{ju} + \sum_{i \leqq j \leqq l=1}^{k} \beta_{ijl} x_{iu} x_{ju} x_{lu} \tag{2.2}$$

or in vector notation as

$$\eta_u = \underline{x}_u \underline{\beta}. \tag{2.3}$$

(For what is to follow, the order of the terms in (2.3) is different from that in (2.2).) If the $(N \times L)$ matrix $\underline{X}$ is defined as

$$\underline{X} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_N \end{bmatrix}, \tag{2.4}$$

where $L = \binom{k+3}{3}$, the number of terms in (2.2), and if $\underline{X}'$ is the transpose of $\underline{X}$, then $N^{-1}(\underline{X}'\underline{X})$ is the moment matrix of a configuration of $N$ points in the $k$-dimensional factor space.

For the configuration of $N$ points, or the design, to be rotatable, $N^{-1}(\underline{X}'\underline{X})$ must satisfy

$$N^{-1}(\underline{X}'\underline{X}) = \begin{bmatrix} \underline{G} & \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ & \lambda_4 \underline{I} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ & & \underline{K}_1 & \underline{0} & \cdots & \underline{0} & \underline{0} \\ & & & \underline{K}_2 & \cdots & \underline{0} & \underline{0} \\ & & & & \ddots & \vdots & \vdots \\ & \text{(symmetric)} & & & & \underline{K}_k & \underline{0} \\ & & & & & & \lambda_6 \underline{I} \end{bmatrix} \tag{2.5}$$

in which the submatrices are defined as follows:

$$\underline{G} = \begin{array}{c} \begin{matrix} \underline{x_0} & \underline{x_1^2} & \underline{x_2^2} & \cdots & \underline{x_k^2} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ & 3\lambda_4 & \lambda_4 & \cdots & \lambda_4 \\ & & 3\lambda_4 & \cdots & \lambda_4 \\ & & & \ddots & \vdots \\ & & & & 3\lambda_4 \end{bmatrix} \end{array} \tag{2.6}$$

$$
(2.7)\qquad K_i \begin{array}{c} \begin{array}{cccccc} x_i & x_i^3 & x_i x_1^2 & x_i x_2^2 & \cdots & x_i x_k^2 \end{array} \\ \begin{bmatrix} 1 & 3\lambda_4 & \lambda_4 & \lambda_4 & \cdots & \lambda_4 \\ & 15\lambda_6 & 3\lambda_6 & 3\lambda_6 & \cdots & 3\lambda_6 \\ & & 3\lambda_6 & \lambda_6 & \cdots & \lambda_6 \\ & & & 3\lambda_6 & \cdots & \lambda_6 \\ & & & & \ddots & \vdots \\ & & & & & 3\lambda_6 \end{bmatrix} \end{array}, \qquad (i = 1, 2, \cdots, k)
$$

$$
(2.8)\qquad \lambda_4 I = \begin{array}{c} \begin{array}{cccc} x_1 x_2 & x_1 x_3 & & x_{k-1} x_k \end{array} \\ \begin{bmatrix} \lambda_4 & 0 & \cdots & 0 \\ & \lambda_4 & \cdots & 0 \\ & & \ddots & \vdots \\ & & & \lambda_4 \end{bmatrix} \end{array},
$$

$$
(2.9)\qquad \lambda_6 I = \begin{array}{c} \begin{array}{cccc} x_1 x_2 x_3 & x_1 x_2 x_4 & \cdots & x_{k-2} x_{k-1} x_k \end{array} \\ \begin{bmatrix} \lambda_6 & 0 & \cdots & 0 \\ & \lambda_6 & \cdots & 0 \\ & & \ddots & \vdots \\ & & & \lambda_6 \end{bmatrix} \end{array}
$$

The headings at the top of the matrices in (2.6) through (2.9) are intended to indicate the form of the elements in the matrices; they are not the vectors of (2.4). The reader will note that the arrangement of the moment matrix (2.5) is different from the arrangement of the second order moment matrix in [1]. (2.5) is written in this form to point out the amount of orthogonality present and to facilitate the calculation of the inverse.

In (2.5), $\mathbf{0}$ denotes a null matrix of appropriate size and in (2.7) the column and row corresponding to $x_i^3$ appears only once and always in the second position.

The constants, $\lambda_4$ and $\lambda_6$, must satisfy the restrictions

$$
(2.10)\qquad \lambda_4 > \frac{k}{k+2}
$$

$$
(2.11)\qquad \lambda_6 > \frac{\lambda_4^2(k+2)}{k+4}
$$

if (2.5) is to be positive definite.

The criterion of rotatability for a third order design is characterized mathematically by equation (2.5) with its attendant restrictions, equations (2.10)

and (2.11). To find a 3rd ORD in $k$ factors, one must discover a set of combinations of factor levels whose moments are those of equation (2.5).

The inverses of the submatrices in (2.5) are

$$G^{-1} = A\begin{bmatrix} b & c & c & \cdots & c \\ & d & e & \cdots & e \\ & & d & \cdots & e \\ & & & \ddots & \vdots \\ & & & & d \end{bmatrix}$$

$$K_i^{-1} = B\begin{bmatrix} f & g & g & g & \cdots & g \\ & h & m & m & \cdots & m \\ & & w & m & \cdots & m \\ & & & w & \cdots & m \\ & & & & \ddots & \vdots \\ & & & & & w \end{bmatrix}$$

$$(\lambda_4 I)^{-1} = I/\lambda_4 \qquad (\lambda_6 I)^{-1} = I/\lambda_6$$

in which

$$b = 2(k+2)\lambda_4^2 \qquad c = -2\lambda_4 \qquad d = (k+1)\lambda_4 - k + 1$$

$$e = 1 - \lambda_4 \qquad f = 6(k+r)\lambda_6$$

$$g = -6\lambda_4 \qquad h = k + 1 - (k-1)\lambda_4^2/\lambda_6 \qquad m = 3\left(\frac{\lambda_4^2}{\lambda_6} - 1\right)$$

$$w = 3[k + 3 - (k+1)\lambda_4^2/\lambda_6]$$

and where $A$ and $B$ are given by

$$1/A = 2\lambda_4[(k+2)\lambda_4 - k] \tag{2.12}$$

$$1/B = 6[(k+4)\lambda_6 - (k+2)\lambda_4^2]. \tag{2.13}$$

**3. Third order rotatable designs in two factors.** Consider an arrangement of $n$ points equally spaced on a circle in a two dimensional factor space. In reference [1], Box and Hunter prove that $n > 2d$ is a sufficient condition for all moments through order $2d$ of the coordinates of these points to be invariant under rotation. That is, $n > 2d$ is sufficient for the arrangement to be rotatable of order $d$. We shall prove the necessity of this condition as well.

We shall use a theorem given by Bose and Carter in reference [3], an earlier version of which was stated by Carter in [4]. Let $(x_{1u}, x_{2u})$ $(u = 1, 2, \cdots, n)$ be the $n$ points of any arrangement $A$ (which may be a design) in the space of $x_1$ and $x_2$. Denote by $\alpha(x_{1u})$, $\alpha(x_{2u})$ the coordinates of the point $(x_{1u}, x_{2u})$

after a rotation about the origin through a fixed angle $\alpha$. From Section 2, it is clear that the arrangement $A$ is rotatable of order $d$ if and only if, for any rotation $\alpha$ performed on all $n$ points of $A$,

$$\sum_{u=1}^{n} \alpha^q(x_{1u})\alpha^r(x_{2u}) = \sum_{u=1}^{n} x_{1u}^q x_{2u}^r \quad \text{for} \quad 0 \leqq q, 0 \leqq r, 0 < q + r \leqq 2d. \tag{3.1}$$

The Bose-Carter theorem proceeds as follows: Let $z_u = x_{1u} + ix_{2u}$ and $\alpha(z_u) = z_u e^{i\alpha}$. Also let $\bar{z}_u$ be the complex conjugate of $z_u$, and $\bar{z}_u e^{-i\alpha}$ the complex conjugate of $\alpha(z_u)$. Put $q + r = p$. Then we may write

$$\sum_{u=1}^{n} x_{1u}^q x_{2u}^r = 2^{-p} i^{-r} \sum_{u=1}^{n} (z_u + \bar{z}_u)^q (z_u - \bar{z}_u)^r = 2^{-p} i^{-r} \sum_{s+t=p} m_{st} \sum_{u=1}^{n} z_u^s \bar{z}_u^t, \tag{3.2}$$

where the $m_{st}$ are sums of combinatorial constants, some of which may be zero. Similarly

$$\sum_{u=1}^{n} \alpha^q(x_{1u})\alpha^r(x_{2u}) = 2^{-p} i^{-r} \sum_{s+t=p} m_{st}\, e^{i\alpha(s-t)} \sum_{u=1}^{n} z_u^s \bar{z}_u^t. \tag{3.3}$$

From (3.2) and (3.3) we see that to satisfy (3.1) it is sufficient that

$$\sum_{u=1}^{n} z_u^s \bar{z}_u^t = 0 \text{ for } 0 \leqq s,\ 0 \leqq t,\ 0 < s + t \leqq 2d \text{ unless } s = t. \tag{3.4}$$

Since

$$z^s \bar{z}^t = (x_1 + ix_2)^s (x_1 - ix_2)^t = \sum_{q+r=s+t} n_{qr} x_1^q x_2^r,$$

then

$$e^{i\alpha(s-t)} \sum_{u=1}^{n} z_u^s \bar{z}_u^t = \sum_{q+r=s+t} n_{qr} \sum_{u=1}^{n} \alpha^q(x_{1u})\alpha^r(x_{2u}).$$

Hence, if the arrangement $A$ is rotatable of order $d$ (i.e., if (3.1) holds), then

$$e^{i\alpha(s-t)} \sum_{u=1}^{n} z_u^s \bar{z}_u^t = \sum_{q+r=s+t} n_{qr} \sum_{u=1}^{n} x_{1u}^q x_{2u}^r$$

is independent of $\alpha$, from which it follows that (3.4) must hold. Thus (3.4) is necessary and sufficient in order that $A$ be rotatable of order $d$. This is a statement of the theorem of Bose and Carter.

Now let the $n$ points $(x_{1u}, x_{2u})$ be points equally spaced on a circle of radius $\rho$. Then we may put

$$x_{1u} = \rho \cos(2\pi v/n),$$
$$x_{2u} = \rho \sin(2\pi v/n)$$

whence

$$z_u = \rho e^{2\pi i v/n},$$
$$\bar{z}_u = \rho e^{-2\pi i v/n}, \qquad v = 0, 1, \cdots, n - 1.$$

The arrangement consisting of these $n$ points is rotatable of order $d$ if and only if

$$(3.5)\quad \sum_{u=1}^{n} z_u^s \bar{z}_u^t = \rho^{s+t} \sum_{v=0}^{n-1} e^{2\pi i v(s-t)/n}$$
$$= 0 \quad \text{for } 0 \leqq s, 0 \leqq t, 0 < s + t \leqq 2d, s \neq t$$

which is a corollary of (3.4). By a well known theorem on the roots of unity $\sum_{v=0}^{n-1} e^{2\pi i v(s-t)/n} = 0$ if and only if $s - t$ is not an integer multiple of $n$. One sees immediately that $s - t$ cannot be an integer multiple of $n$ if $s + t < n$ and that $s - t$ will be such a multiple for some $s$ and $t$ if $s + t \geqq n$. Since (3.5) should be satisfied for any non-negative $s, t$ with $0 < s + t \leqq 2d$, then $n > 2d$ is necessary and sufficient for equally spaced points on a circle to be rotatable of order $d$.

Equation (2.5), then, if satisfied by $n > 6$ points equally spaced on a circle. But it may be verified that for this arrangement

$$\lambda_4 = \frac{n \sum_u x_{1u}^2 x_{2u}^2}{\left[\sum_u x_{iu}^2\right]^2} = \frac{n\rho^4 n/8}{(\rho^2 n/2)^2} = \frac{1}{2}, \qquad i = 1, 2,$$

which does not satisfy (2.10). Therefore, these points do not constitute a rotatable design. If $n_1$ is the number of points on the circle and $n_2$ points are added at the center, $\lambda_4$ becomes

$$\lambda_4 = \frac{(n_1 + n_2)\rho^4 n_1/8}{(\rho^2 n_1/2)^2} = \frac{1}{2}\left[1 + \frac{n_2}{n_1}\right] > \frac{1}{2}$$

which satisfies (2.10), but then

$$\lambda_6 = \frac{(n_1 + n_2)^2 \rho^6 n_1/16}{3(\rho^2 n_1/2)^3} = \frac{1}{6}\left[1 + \frac{n_2}{n_1}\right]^2 = \frac{2}{3}\lambda_4^2$$

and (2.11) is not satisfied. Equally spaced points on a circle with additional points at the center, then, do not constitute a 3rd ORD.

Now consider an arrangement of $N$ points on two concentric circles with $n_1$ points equally spaced on a circle of radius $\rho_1$ and $n_2$ points equally spaced on a circle of radius $\rho_2$, where $n_1 + n_2 = N$, $\rho_1 \neq \rho_2$, $\rho_1 > 0$, $\rho_2 > 0$. We shall prove that the arrangement consisting of these $n_1 + n_2$ points is rotatable of order $d$ if and only if both $n_1 > 2d$ and $n_2 > 2d$.

In the same manner as before take the first $n_1$ points as

$$\rho_1 e^{2\pi i v/n_1} \; (v = 0, 1, \cdots, n_1 - 1).$$

To allow the second $n_2$ points to take any orientation with respect to the $n_1$ points, take them as $e^{i\beta}\rho_2 e^{2\pi i v/n_2}$ $(v = 0, 1, \cdots, n_2 - 1)$. The arrangement is rotatable of order $d$ if and only if

$$(3.6)\qquad \rho_1^{s+t} \sum_{v=0}^{n_1-1} e^{2\pi i v(s-t)/n_1} + e^{i(s-t)\beta}\rho_2^{s+t} \sum_{v=0}^{n_2-1} e^{2\pi i v(s-t)/n_2} = 0$$

for $0 \leqq s$, $0 \leqq t$, with $0 < s + t \leqq 2d$ unless $s = t$. But the sums in (3.6) are 0 or $n_1$ and 0 or $n_2$ respectively. Hence (3.6) holds if and only if both sums are zero. In order that this be true we know that $n_1 > 2d$ and $n_2 > 2d$ is necessary and sufficient.

It is easily shown that this type of arrangement provides a 3rd ORD if $n_1$, $n_2 > 6$. For then

$$\lambda_4 = \frac{N(n_1 \rho_1^4 + n_2 \rho_2^4)/8}{[n_1 \rho_1^2 + n_2 \rho_2^2]^2/4} = \frac{1}{2} \frac{n_1^2 \rho_1^4 + n_2^2 \rho_2^4 + n_1 n_2(\rho_1^4 + \rho_2^4)}{n_1^2 \rho_1^4 + n_2^2 \rho_2^4 + 2n_1 n_2 \rho_1^2 \rho_2^2},$$

and since $\rho_1^4 + \rho_2^4 > 2\rho_1^2\rho_2^2$, for $\rho_1 \neq \rho_2$, $\lambda_4 > \frac{1}{2}$. Therefore, (2.10) is satisfied. Similarly

$$\lambda_6 = \frac{N^2}{3} \frac{(n_1 \rho_1^6 + n_2 \rho_2^6)/16}{[n_1 \rho_1^2 + n_2 \rho_2^2]^3/8} > \frac{2}{3} \lambda_4^2$$

since

$$\frac{\lambda_6}{\lambda_4^2} = \frac{2}{3} \frac{n_1^2 \rho_1^8 + n_2^2 \rho_2^8 + n_1 n_2 \rho_1^2 \rho_2^2(\rho_1^4 + \rho_2^4)}{n_1^2 \rho_1^8 + n_2^2 \rho_2^8 + n_1 n_2 \rho_1^2 \rho_2^2(2\rho_1^2 \rho_2^2)}$$

which is greater than $\frac{2}{3}$ for $\rho_1 \neq \rho_2$ and so (2.11) is satisfied, also.

Thus, it has been shown that a simple class of 3rd ORDs in two factors exists. This class consists of designs which have seven or more points equally spaced on each of two concentric circles. Each of the circles may be rotated independently of the other and therefore there are an infinite number of configurations possible for designs with a given $n_1$ and $n_2$. Since points located at the center of the circles do not disturb the moment properties of the configuration, these may be added at will to achieve variations in the parameters $\lambda_4$ and $\lambda_6$.

**4. Sequential third order rotatable designs in two factors.** A 3rd ORD of the type described in the previous section may be performed in two "blocks." By judicious selection of $\rho_1$ and $\rho_2$, the radii of the two circles, the coefficients in the Taylor series expansion may be estimated independently of the block effects. If one block of points is a complete 2nd ORD and the second block consists of additional points necessary to make the whole a 3rd ORD, the design may be called sequential, in that an experimenter need not perform the second block of points if he feels the first block has given him an adequate approximation to the phenomenon.

Suppose the first block consists of seven or more points equally spaced on a circle with some points at the center. This allows the estimation of polynomial coefficients up to and including the second order. Now add a second block consisting of seven or more points equally spaced on a circle of different radius from the first. Let $n_1$ be the number of points in the first block and $n_2$ the number of points in the second block. Let $\delta_1$ be the effect of the first block, $\delta_2$ the effect of the second block, and let $Z_{wu} = 1$ if the $u$th observation occurs in the $w$th

block, $w = 1, 2$, and $Z_{wu} = 0$ otherwise. Then, the expectation of the $u$th observation can be written

$$\eta_u = \beta_0 + \sum_i \beta_i x_{iu} + \sum_i \sum_j \beta_{ij} x_{iu} x_{ju}$$

$$(4.1) \qquad + \sum_i \sum \sum \beta_{ijl} x_{iu} x_{ju} x_{lu} + \sum_w \delta_w (Z_{wu} - \bar{Z}_{wu})$$

in which $\bar{Z}_w = \sum_u Z_{wu}/N$, and $N = n_1 + n_2$.

If the estimates of the block effects are to be independent of the estimates of the polynomial coefficients, it is required that

$$(4.2) \qquad \sum_u (Z_{wu} - \bar{Z}_w) = 0$$

$$(4.3) \qquad \sum_u (Z_{wu} - \bar{Z}_w) x_{iu} = 0$$

$$(4.4) \qquad \sum_u (Z_{wu} - \bar{Z}_w) x_{iu} x_{ju} = 0$$

$$(4.5) \qquad \sum_u (Z_{wu} - \bar{Z}_w) x_{iu} x_{ju} x_{lu} = 0$$

for $w = 1, 2$ and $i, j, l = 1, 2$. (4.2) is satisfied by the definition of $\bar{Z}_w$, while (4.3), (4.4) and (4.5) are satisfied with one exception, by the fact that $Z_{wu} - \bar{Z}_w$ is constant within blocks and each block contains a rotatable arrangement of points. The exception is in (4.4) when $i = j$. For this case, if $n_{01}$ = the number of points at the center in the first block, and $n_{02}$ = the number of points at the center in the second block, (4.4) becomes

$$\left[1 - \frac{n_1}{N}\right][n_1 - n_{01}] \frac{\rho_1^2}{2} + \left[\frac{-n_1}{N}\right][n_2 - n_{02}] \frac{\rho_2^2}{2} = 0$$

or

$$(4.6) \qquad \frac{\rho_2^2}{\rho_1^2} = \frac{n_2(n_1 - n_{01})}{n_1(n_2 - n_{02})}.$$

Therefore, by selecting $\rho_2$, the radius of the circle in the second block, in accordance with (4.6) the experiment may be performed sequentially and estimates of polynomial coefficients will be free of block effects. It is interesting to note that (4.6) is independent of the number of points in the second block, if $n_{02} = 0$, it being required only that $n_2 > 6$. A 3rd ORD with these blocking properties is not possible, however, if $n_2 n_{01} = n_1 n_{02}$.

The 3rd ORD may be sequentialized in three stages with a total of either three or four blocks: Block I would consist of $n_1/2$ points of which $n_{01}/2$ are central points and such that $(n_1 - n_{01})/2$ is an integer greater than or equal to 4. The $(n_1 - n_{01})/2$ points would be equally spaced on a circle of radius $\rho_1$, and the $(n_1 - n_{01})/2$ points would constitute a 1st ORD. Block II would be identical with Block I with the points superposed so that Blocks I and II jointly would have $n_1 - n_{01}$ points equally spaced on a circle of radius $\rho_1$ and $n_{01}$ points at the center and thus would form a complete 2nd ORD.

The third stage, Block III, would consist of $n_3 - n_{03}$ points, greater than 6, equally spaced on a circle of radius $\rho_3$ (where $\rho_3$ is determined from (4.6)) and $n_{03}$ points at the center, in a three block design. Blocks I, II, and III would make up a complete 3rd ORD.

If the experiment were to be sequentialized in three stages and four blocks, the third stage would be constructed of two blocks similar to Blocks I and II, but with radius $\rho_3$, and with the possibility of no central points.

**5. Third order rotatable designs in three factors (non-sequential).** A 3rd ORD in three factors may be formed from the points at the vertices of a cube, two octahedra of different radii, and a cuboctahedron, all oriented symmetrically to one another. The coordinates of the points of the cube can be represented by all possible permutations of the elements of the vector, $(\pm a, \pm a, \pm a)$; of one octahedron by the permutations of the elements of $(\pm 1.82969a, 0, 0)$; of the other octahedron by the permutations of the elements of $(\pm 1.16343a, 0, 0)$; and of the cuboctahedron by the 12 permutations of the elements of $(\pm a2^{1/2}, \pm a2^{1/2}, 0)$. The value of $a$ is the scaling factor chosen so that $\sum_{u=1}^{N} x_{iu}^2 = N$, the total number of points. The constants, 1.82969, 1.16343, and $2^{1/2}$ are those which will satisfy the moment requirements inherent in equations (2.6) and (2.7) for this composite configuration. The parameters for this design are given below for various numbers of points added at the center of the design. Also given are the values $(5/7)\lambda_4^2$, which, in accordance with (2.11), must always be exceeded by $\lambda_6$.

| $N$ | No. of Center Points | $\lambda_4$ | $\lambda_6$ | $5\lambda_4^2/7$ |
|---|---|---|---|---|
| 32 | 0 | .638 | .300 | .291 |
| 33 | 1 | .658 | .319 | .309 |
| 34 | 2 | .678 | .339 | .328 |
| 35 | 3 | .698 | .359 | .348 |
| 36 | 4 | .718 | .380 | .368 |
| 37 | 5 | .738 | .402 | .389 |
| 38 | 6 | .758 | .423 | .410 |
| 39 | 7 | .778 | .446 | .432 |
| 40 | 8 | .798 | .469 | .455 |

Another 3rd ORD of the non-sequential type can be formed by orienting an icosahedron of radius $a$ symmetrically with respect to a dodecahedron of radius $1.11236224a$, with or without central points. But with 0 to 8 central points $\lambda_6 - (5/7)\lambda_4^2$ is never greater than 0.000061, so that this design could not be recommended.

**6. Third order rotatable designs in three factors (sequential).** Of greater interest than a 3rd ORD *per se* is the 3rd ORD which can be performed sequentially, and particularly those sequential designs in three factors which may be extended to higher dimensions.

Consider the sequential design as being performed in two parts: the first part to be a 2nd ORD and the second part a set of points which, when added to the first part, makes a 3rd ORD. If the second order moment properties are to be preserved after the addition of the second set of points, it is obvious than both parts of the design must be complete 2nd ORDs in themselves.

For the first of these 3 dimensional sequential designs consider the design whose initial portion is the cube + octahedron configuration with points at the center. By adding to this, the points of a truncated cube and of another octahedron, a 3rd ORD results. This design would not be recommended in practice because, like the icosahedron-dodecahedron design of the previous section, the resultant matrix of normal equations is poorly conditioned. That is, although the inequalities (2.10) and (2.11) are satisfied, (2.11) is very close to an equality. Consequently, the linear and cubic coefficients are very poorly estimated. The design is presented here because it provides a basis for a more useful design which follows.

The coordinates of a truncated cube in 3 dimensions can be written as all 24 permutations of the elements of the vector, $(\pm c, \pm d, \pm d)$, where the radius of the figure is given by $\rho^2 = c^2 + 2d^2$ and where $c$ is a measure of the amount of truncation. For example, if $c = d$ the figure is not truncated at all and the 24 points make up a triply replicated cube. If $c = 0$, the truncation is extreme and the figure becomes a doubly replicated cuboctahedron. It can be shown that if

$$c = \frac{5 + 2\sqrt{10}}{15} \rho^2$$

the 24 points constitute a 2nd ORD, but this value of $c$ is not helpful in constructing a 3rd ORD.

For the first portion of this sequential design let the cube have radius $\rho_1$ and an octahedron have radius

$$\rho_2 = \frac{2^{3/4}}{\sqrt{3}} \rho_1 .$$

Box and Hunter [1] show that this arrangement of 14 points comprises a 2nd ORD. To this, as the second portion of the sequential design, add a truncated cube of radius $\rho_3$ and coordinates, $(\pm c, \pm d, \pm d)$ and another octahedron with coordinates $(\pm \rho_4, 0, 0)$. Then it can be shown that if $\rho_1 = \sqrt{3}$, $\rho_2 = 2^{3/4}$, $\rho_3 = 1.657765$, $\rho_4 = 1.705945$, $c = 0.184388$, $d = 1.164944$, a 3rd ORD results. But for 0 to 10 central points $\lambda_6 - 5\lambda_4^2/7$ is never larger than .0005.

However, the difficulty of the ill-conditioned matrix may be avoided by modifying the design slightly. This is accomplished by using a cube and "doubled octahedron," instead of a cube and octahedron in the first stage of experimentation. By doubled octahedron is meant *two* experimental points at each vertex of an octahedron. Let $\rho_1$, $\rho_2$, $\rho_3$ and $\rho_4$ designate the radii of a cube, doubled octahedron, truncated cube and octahedron respectively. If the first portion, consisting of the cube and doubled octahedron is to be second order rotatable, then

$\rho_2 = \rho_1\sqrt{2/3}$. The second portion of the design consisting of the truncated cube and the other octahedron must have dimensions satisfying the equations

$$\rho_4^4 = \rho_3^4 + 10c^2\rho_3^2 - 15c^4 \tag{6.1}$$

$$2\rho_4^6 = -21c^6 + 9c^4\rho_3^2 + c^2\rho_3^4 + 3\rho_3^6 \tag{6.2}$$

$$\tfrac{16}{27}\rho_1^6 = \rho_3^6 - 19c^2\rho_3^4 + 39c^4\rho_3^2 - 21c^6 \tag{6.3}$$

which, in turn, will satisfy equation (2.5). If $\rho_1$ again equals $\sqrt{3}$ and hence $\rho_2 = \sqrt{2}$, then an admissable solution of (6.1), (6.2) and (6.3) is $\rho_3 = 1.851208$, $\rho_4 = 1.985406$, $c = 0.341564$. The coordinates of the resultant design are
(for the first stage of experimentation)
the 8 permutations of $(\pm a, \pm a, \pm a)$,
the 6 permutations of $(\pm a\sqrt{2}, 0, 0)$,
the 6 permutations of $(\pm a\sqrt{2}, 0, 0)$ again,
and, if desired, points with coordinates $(0, 0, 0)$;
(for the second stage of experimentation)
the 24 permutations of $(\pm.341564a, \pm 1.286527a, \pm 1.286527a)$,
the 6 permutations of $(\pm 1.985406a, 0, 0)$,
and, if desired, points with coordinates $(0, 0, 0)$,
where, again, $a$ is chosen so that $\sum_{u=1}^{N} x_{iu}^2 = N$.

The values of the parameters for number of central points to 10 are:

| $N$ | Number of Central Points | $\lambda_4$ | $\lambda_6$ | $5\lambda_4^2/7$ |
|---|---|---|---|---|
| 50 | 0 | .6271 | .2902 | .2809 |
| 51 | 1 | .6396 | .3019 | .2922 |
| 52 | 2 | .6522 | .3139 | .3038 |
| 53 | 3 | .6647 | .3261 | .3156 |
| 54 | 4 | .6773 | .3385 | .3277 |
| 55 | 5 | .6898 | .3511 | .3399 |
| 56 | 6 | .7023 | .3640 | .3523 |
| 57 | 7 | .7149 | .3771 | .3651 |
| 58 | 8 | .7274 | .3905 | .3779 |
| 59 | 9 | .7400 | .4041 | .3911 |
| 60 | 10 | .7525 | .4179 | .4045 |

Block coefficients may be introduced into the model with this design also. Following Section 4, $n_1 - n_{01} = 20$ and $n_2 - n_{02} = 30$, so the equation which expresses the condition for orthogonal blocking is

$$(30 + n_{02}) \sum_{u=1}^{n_1} x_{iu}^2 = (20 + n_{01}) \sum_{u=n_1+1}^{N} x_{iu}^2$$

or

$$n_{02} = 2.206n_{01} + 14.124 \tag{6.4}$$

Applying equation (6.4), the following table of design numbers for approximately orthogonal blocking results.

| $n_{01}$ | $n_{02}$ | $n_1$ | $n_2$ | $N$ |
|---|---|---|---|---|
| 0 | 14 | 20 | 44 | 64 |
| 1 | 16 | 21 | 46 | 67 |
| 2 | 19 | 22 | 49 | 71 |
| 3 | 21 | 23 | 51 | 74 |
| 4 | 23 | 24 | 53 | 77 |
| 5 | 25 | 25 | 55 | 80 |
| 6 | 27 | 26 | 57 | 83 |
| 7 | 30 | 27 | 60 | 87 |

**7. Third order rotatable designs in more than three factors.** As is well known, only the analogues of the tetrahedron, the octahedron and the cube exist, as regular figures in more than four dimensions. The latter two were used successfully by Box and Hunter [1] in their development of 2nd ORDs in the higher dimensional factor spaces. In Sections 5 and 6 some semi-regular figures were described which provided 3rd ORDs for three dimensions. Of these semi-regular figures, the truncated cube was of particular interest in that it provided a basis for the construction of three dimensional sequential 3rd ORDs.

The higher dimensional analogue of the truncated cube is not easy to identify. The obvious extension from three dimensions to $k$ dimensions would be the figure whose coordinates are the permutations of the elements of $(\pm c, \pm d, \cdots, \pm d)$, there being $k - 1$ elements $\pm d$. Call this "truncated cube (1)." A less obvious, but nevertheless reasonable extension to $k$ dimensions is the figure whose coordinates are the permutations of $(\pm c, \pm c, \cdots, \pm c, \pm d, \pm d, \cdots, \pm d)$ with, say, $r$ elements $\pm c$ and $k - r$ elements $\pm d$, and with $1 < r < (k + 1)/2$. Let this figure be called "truncated cube $(r)$."

From the point of view of economy in the number of experiments, "truncated cube (1)" would be preferred as a part of a design since it contains fewer points. The number of points in "truncated cube (1)" is $k2^k$. The number of points in "truncated cube $(r)$" is $\binom{k}{r}2^k$, which is always larger for $1 < r < k - 1$. Unfortunately "truncated cube (1)" cannot be used with the "cube" and "octahedra" to form sequential 3rd ORDs for $k > 3$. This can be shown as follows.

Consider the configuration made up of the $k$-dimensional "truncated cube (1)" and a $k$-dimensional "octahedron" of radius $\rho$. For a sequential design of the type described in Section 6, this composite configuration would comprise the second stage of experimentation and therefore the coordinates must satisfy the requirement (in addition to the requirements satisfied by the symmetry of the configuration),

$$\sum_u x_{iu}^4 = 3 \sum_u x_{iu}^2 x_{ju}^2 \qquad (i \neq j = 1, 2, \cdots, k). \tag{7.1}$$

For this configuration

$$\sum_u x_{iu}^4 = 2^k[c^4 + (k-1)\, d^4] + 2\rho^4$$

$$\sum_u x_{iu}^2 x_{ju}^2 = 2^k d^2[2c^2 + (k-2)\, d^2].$$

Substitution of the above into (7.1), requires

$$c^2 = 3\, d^2 \pm [(4+2k)\, d^4 - 2^{1-k}\rho^4]^{\frac{1}{2}}. \tag{7.2}$$

The configuration for the second stage when combined with the first stage configuration consisting of a $k$-dimensional "cube" of radius $(k)^{\frac{1}{2}}$ and $k$-dimensional "octahedron" of radius $(2)^{k/4}$ must satisfy the condition

$$\sum_u x_{iu}^4 x_{ju}^2 = 3 \sum_u x_{iu}^2 x_{ju}^2 x_{lu}^2, \quad (i \neq j \neq l = 1, 2, \cdots, k). \tag{7.3}$$

For the combined configurations (7.3) requires

$$c^2 = \frac{4d^4 \pm [(9+2k)d^8 + 2d^2]^{\frac{1}{2}}}{d^2} \tag{7.4}$$

The minus sign in (7.4) will give a negative $c^2$ (with $k > 3$) and therefore must be disregarded. Equating (7.2) to (7.4) (with the plus sign) and simplifying gives

$$3d^6 + d^2[2^{-k}\rho^4 + \sqrt{(9+2k)\, d^8 + 2d^2}] + 1 = 0, \tag{7.5}$$

which is impossible since each term on the left of (7.5) must be a positive quantity. Therefore, "truncated cube (1)" cannot be used in a 3rd ORD of this form if $k$ is greater than 3.

With $k = 4$ a sequential 3rd ORD was discovered. The first stage of the design consists of a four dimensional "cube" of radius 2 and a 4-dimensional "octahedron," also of radius 2. (Actually, this is a 4-dimensional regular figure of 24 points.) The second stage is comprised of a 4-dimensional "truncated cube (2)" with coordinates $(\pm c, \pm c, \pm d, \pm d)$ and another 4-dimensional "octahedron" of radius $\rho$. To satisfy equation (2.5), we must have $c = 1.200919$, $d = .256303$, $\rho = 1.736604$. This design contains 16 points on the "cube," 8 points on the first "octahedron," 96 points on the "truncated cube" and 8 points on the second "octahedron" for a total of 128 points without center points. The design parameters are $\lambda_4 = .676$ and $\lambda_6 = .349$. The coordinates of experimental points for this design are

(for the first stage)

the 16 permutations of $(\pm a, \pm a, \pm a, \pm a)$,
the 8 permutations of $(\pm 2a, 0, 0, 0)$,
and, if desired, central points $(0, 0, 0, 0)$;

(for the second stage)

the 96 permutations of $(\pm 1.200919a, \pm 1.200919a, \pm .256303, \pm .256303a)$,
the 8 permutations of $(\pm 1.736604a, 0, 0, 0)$,
and, if desired, central points $(0, 0, 0, 0)$,

with $a$ such that $\sum_{u=1}^{N} x_{iu}^2 = N$.

For approximately orthogonal blocking of the two stages of experimentation, the number of center points in each block, $n_{01}$ and $n_{02}$, is shown in the following table. The total number of points and the design parameters are also given.

| $n_{01}$ | $n_{02}$ | $n_1$ | $n_2$ | $N$ | $\lambda_6$ | $\lambda_4$ | $6\lambda_4^2/8$ |
|---|---|---|---|---|---|---|---|
| 8 | 0 | 32 | 104 | 136 | .719 | .394 | .388 |
| 9 | 4 | 33 | 108 | 141 | .745 | .423 | .416 |
| 10 | 7 | 34 | 111 | 145 | .766 | .447 | .440 |
| 11 | 10 | 35 | 114 | 149 | .787 | .472 | .465 |

The relationship, $\lambda_6 > 6\lambda_4^2/8$, appears to be sufficiently well satisfied so that no investigation of the design utilizing a doubled octahedron (and 8 additional points) was made.

No attempt was made to extend the concept of 3rd ORDs to more than four dimensions, chiefly because the approach pursued in this paper required the use of an excessive number of experimental points. Investigations were made, following this approach, only of the sequential type of rotatable design because this is the type which seems likely to be most useful to an experimenter.

Considerable savings were demonstrated by Box and Hunter in the case of 2nd ORDs by the use of fractional replication for $k > 4$. With $k$ equal to five or more the second order coefficients are confounded only with third and higher order effects when fractional replication is used. But for third order coefficients to be confounded only with fourth and higher effects, the dimensionality must be at least seven in order to make use of fractional replication. If a half replicate of a 7-dimensional design of the type described in the preceding section were possible it would require at least 1,436 experimental points.

If a full replicate, 5-dimensional design of this type were possible, 372 points would be required. The same design in six factors would require 1,048. Third order rotatable designs derived from figures which are symmetrical in all $k$-dimensions would appear to be impractical for $k > 4$.

**8. Summary and conclusions.** This paper is concerned with extending the criterion of rotatability, as advanced by Box and Hunter [1], to experimental designs for estimating response surfaces by third order polynomial equations. The method of attack has been to examine combinations of regular and semi-regular geometrical figures and find those combinations whose coordinate points satisfy the moment properties, to order six, of spherical distributions. Designs with these properties and the attendant restrictions were shown by Box and Hunter to have spherical variance contours when the polynomial coefficients were estimated by the method of least squares.

It was found that 3rd ORDs in two factors could be attained by locating seven or more experimental points equally spaced on each of two concentric circles of different non-zero radii. Also it was shown that certain rotatable designs in two factors can be performed in two stages, so that second order polynomial coefficients can be estimated after the first stage and third order poly-

nomial coefficients after the second stage. By choosing the radii of the two circles in the proper ratio it is possible to obtain estimates of the polynomial coefficients which are independent of "block" effects due to running the experiments in two stages. Such designs were termed sequential 3rd ORDs.

In three factors, 3rd ORDs were presented which consisted of composites of cubes, truncated cubes, octahedra, cuboctahedra, icosahedra and dodecahedra. Two of these designs in three factors were constructed so that they might be performed sequentially.

One sequential 3rd ORD in four factors was also presented. This design has as its experimental points the vertices of the 4-dimensional analogues of a cube, a truncated cube and two octahedra of different dimensions.

**9. Acknowledgement.** The authors express their gratitude to the referee who called their attention to the theorem of Bose and Carter and suggested the proofs given in Section 3. As a result, Section 3 was shortened and considerably improved.

## REFERENCES

[1] G. E. P. Box and J. S. Hunter, "Multi-factor Experimental Designs for Exploring Response Surfaces," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 195–241.

[2] D. A. Gardiner, "Some Third Order Rotatable Designs," Ph.D. Thesis, North Carolina State College, 1956.

[3] R. C. Bose and R. L. Carter, "Complex Representation in the Construction of Rotatable Designs," *Ann. Math. Stat.*, Vol. 30, (1959), pp. 771–780.

[4] Carter, Richard Leston, "New Designs for the Exploration of Response Surfaces," *Technical Report No.* 172, 1957, Institute of Statistics, University of North Carolina.

# SECOND ORDER ROTATABLE DESIGNS IN THREE DIMENSIONS[1]

By R. C. Bose and Norman R. Draper[2]

*University of North Carolina*

**0. Summary.** The technique of fitting a response surface is one widely used (especially in the chemical industry) to aid in the statistical analysis of experimental work in which the "yield" of a product depends, in some unknown fashion, on one or more controllable variables. Before the details of such an analysis can be carried out, experiments must be performed at predetermined levels of the controllable factors, i.e., an experimental design must be selected prior to experimentation. Box and Hunter [3] suggested designs of a certain type, which they called rotatable, as being suitable for such experimentation. Very few of these designs were then known. Since that time the work of R. L. Carter [6] has provided many new second order rotatable designs in two factors. However, additional methods were needed which would provide both second and third order designs in three and more factors. The present work represents an attempt to meet, in part, this need. New construction methods for obtaining rotatable designs of second order in three dimensions are here presented. By use of these methods various infinite classes of designs are obtained, and it may be shown that all the rotatable designs previously known can be derived as special cases of these infinite classes. Also derived is an infinite class of second order rotatable designs which contain only 16 points; only two particular designs contain fewer points.

**1. Introduction.** A great deal of information is now available about the theory of response surfaces and the use of rotatable designs. Such information may be found in papers by Box [1], [2], Box and Wilson [5], Box and Hunter [3], [4] and the Ph.D. dissertation of Carter [6]. The paper [3] by Box and Hunter provides the necessary background for the present work, and a discussion of polynomial approximation and of the desirability of rotatable designs will be found therein. We shall be concerned here with second order rotatable designs in three controllable factors and we shall assume that the measurements of the factors have been coded, permitting the use of Cartesian axes in three dimensional space to describe an experimental design.

Suppose, in an experimental investigation with $k$ factors, $N$ (not necessarily

Received June 27, 1958; revised July 6, 1959.

[1] This work was supported in part by the Office of Naval Research under contract for research in probability and statistics at Chapel Hill; in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(600)-83; and in part by a Bell Telephone Graduate Fellowship awarded to the second author, who gratefully acknowledges his indebtedness. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now with Imperial Chemical Industries Limited, England.

distinct) combinations of levels are employed. Thus the group of $N$ experiments which arises can be described by the $N$ points in $k$ dimensions

$$(1.1) \qquad (x_{1u}, x_{2u}, \cdots, x_{ku}), \qquad u = 1, 2, \cdots, N;$$

where, in the $u$th experiment, factor $t$ is at level $x_{tu}$.

The set of points (1.1) is said to form a *rotatable arrangement* of the second order in $k$ factors if the following conditions are satisfied:

$$(1.2) \qquad \begin{aligned} &\sum_u x_{1u}^2 = \sum_u x_{2u}^2 = \cdots = \sum_u x_{ku}^2 = \lambda_2 N, \\ &\sum_u x_{1u}^4 = \sum_u x_{2u}^4 = \cdots = \sum_u x_{ku}^4 = 3\sum_u x_{iu}^2 x_{ju}^2 = 3\lambda_4 N, \quad (i \neq j) \end{aligned}$$

and all other sums of powers and products up to and including order four are zero, where all summations are over $u = 1$ to $u = N$. The set (1.1) is said to form a *rotatable design* of second order if the conditions (1.2) are satisfied *and* a certain matrix used in a subsequent least squares estimation is non-singular. Box and Hunter [3] show that the necessary and sufficient condition for this to be so is

$$(1.3) \qquad \lambda_4/\lambda_2^2 > k/(k+2),$$

a condition which may always be satisfied merely by theaddition of points at the center (0, 0, 0) of the design. Equality in (1.3) is attained when all the design points lie on a $k$-dimensional sphere, and it is impossible for the inequality in (1.3) to be reversed under any circumstances.

When presenting a rotatable design, it is customary to "scale" it. By this it is meant that the scale of the coded controllable variables is chosen in such a way that $\lambda_2 = 1$. The reason for this is as follows. Given a second order design which satisfies the conditions (1.2) with a *specified* value of $\lambda_4/\lambda_2^2$, there are an infinite number of values possible for $\lambda_2 > 0$. Since these designs can be derived one from the other merely by change of scale, we do not regard them as different. Thus the scaling condition $\lambda_2 = 1$ fixes a *particular* design and enables better comparison between two designs with different values of $\lambda_4/\lambda_2^2$.

**2. A transformation group in three dimensions and its generated point sets.** We shall define certain transformations applied to points in three dimensions. Let $W(x, y, z) = (y, z, x)$. Then $W^2(x, y, z) = (z, x, y)$ and $W^3(x, y, z) = (x, y, z)$. Thus $W$, $W^2$ and $W^3 = I$ form a cyclical group of order 3. Further let $R_1(x, y, z) = (-x, y, z)$, $R_2(x, y, z) = (x, -y, z)$, $R_3(x, y, z) = (x, y, -z)$.

The four transformations represented by $W$, $R_1$, and $R_2$ and $R_3$ generate a group $G$ of transformations of order 24 with elements

$$(2.1) \qquad \begin{aligned} &W^j, W^jR_1, W^jR_2, W^jR_3, W^jR_2R_3, \\ &W^jR_3R_1, W^jR_1R_2, W^jR_1R_2R_3 \qquad (j = 1, 2, 3). \end{aligned}$$

It is easily seen that all the 24 elements in (2.1) are distinct. While $R_1$, $R_2$ and $R_3$ commute, $W^j$ and $R_i$ do not ($j = 1, 2$; $i = 1, 2, 3$).

A group table may be constructed, employing the identities

$$W^3 = R_1^2 = R_2^2 = R_3^2 = I \tag{2.2}$$

and identities of the type $WR_1 = R_3W$, to verify the statements above. Because of the size of the group the table will not be reproduced here.

Given a general point $(x, y, z)$ in three dimensions, we may apply to it all the transformations of the group $G$. In this way we obtain a set of 24 points with coordinates

$$(\pm x, \pm y, \pm z), \qquad (\pm y, \pm z, \pm x), \qquad (\pm z, \pm x, \pm y). \tag{2.3}$$

We shall denote this set by

$$G(x, y, z). \tag{2.4}$$

Note that if $(l, m, n)$ denotes any other point of the set, $G(x, y, z) = G(l, m, n)$, i.e., any point of the set, when operated on by $G$, will give rise to the same set. The set $G(x, y, z)$ satisfies all the moment conditions (1.2) except

$$\sum_{u=1}^{N} x_{iu}^4 = 3\sum_{u=1}^{N} x_{iu}^2 x_{ju}^2 \qquad (i \neq j), \quad (i, j = 1, 2, 3). \tag{2.5}$$

We now define a function $K(x, y, z)$ of the point $(x, y, z)$ as

$$K(x, y, z) = \tfrac{1}{8}(x^4 + y^4 + z^4 - 3y^2z^2 - 3z^2x^2 - 3x^2y^2). \tag{2.6}$$

This function is constant for all of the 24 points of $G(x, y, z)$. Furthermore, if it has the value zero, then $G(x, y, z)$ is a rotatable arrangement since the outstanding condition (2.5) becomes satisfied. Let

$$x^2 = sz^2, \qquad y^2 = tz^2. \tag{2.7}$$

Then, if $K(x, y, z)$ is zero and $z \neq 0$,

$$t^2 - 3t(s + 1) + (s^2 - 3s + 1) = 0. \tag{2.8}$$

This is the equation of a hyperbola. If the point $(s, t)$ lies on the hyperbola and also in the first quadrant, $G(x, y, z)$ is a rotatable arrangement. Fig. 1 shows points $(s, t)$ for which this is true. There is complete symmetry about the line $s = t$. The value of $s$ at the points $P_1$ and $P_2$, where the hyperbola intersects the line $t = 0$, is $(3 - \sqrt{5})/2$ and $(3 + \sqrt{5})/2$, respectively. If we solve for $t$ in terms of $s$, we obtain

$$t = \tfrac{1}{2}[3(s + 1) \pm \sqrt{5(s^2 + 6s + 1)}]. \tag{2.9}$$

This yields two non-negative solutions if $s^2 - 3s + 1 > 0$, which implies $s \geqq (3 + \sqrt{5})/2$ or $0 \leqq s \leqq (3 - \sqrt{5})/2$. Otherwise there is only one positive solution for each value of $s \geqq 0$. The reason for this is clear from Fig. 1.

The point set $G(x, y, z)$ is clearly spherical, and thus equality will be attained in the non-singularity condition (1.3) unless additional points are added at the center to form the design. If $n_0$ center points are added, $N = 24 + n_0$, and $\lambda_2 N = 8(x^2 + y^2 + z^2) = 8(s + t + 1)z^2$.

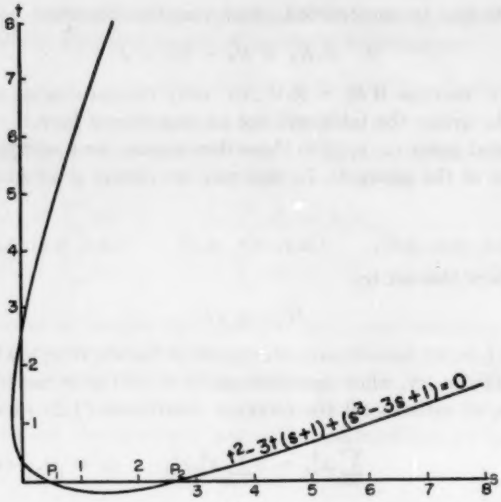


FIG. 1

Thus if we apply the scaling condition $\lambda_2 = 1$,

$$z^2 = N/8(s + t + 1), \tag{2.10}$$

and we have an infinite class of second order designs which depends on one parameter $s$. For if $s \geqq 0$ is specified,

$$\begin{aligned} t &= \tfrac{1}{2}[3(s+1) \pm \sqrt{5(s^2 + 6s + 1)}], \qquad (t \geqq 0 \text{ only}), \\ z &= [N/8(s + t + 1)]^{\frac{1}{2}}, \qquad y = t^{\frac{1}{2}}z, \qquad x = s^{\frac{1}{2}}z, \end{aligned} \tag{2.11}$$

and all design points are fixed. Each non-negative $s$ gives rise to one or two designs according as (2.9) yields one or two non-negative values of $t$. For this class, $\lambda_4/\lambda_2^2 = 8(x^2y^2 + y^2z^2 + z^2x^2)/N = 8(st + s + t)z^4/N$. Consider the special case $s = t = \sqrt{10} - 3$. We then have $x = y = (N - 8z^2)/16$, $z = [(5 + 2\sqrt{10})N/120]^{\frac{1}{2}}$. This is the design referred to as the truncated cube by Gardiner, Grandage and Hader ([9], Sec. 6, Par. 4).

Let us now suppose that $K(x, y, z) \neq 0$ for the points of the set $G(x, y, z)$. We shall define $\sum K(x, y, z)$ over a point set $S$ to be the *excess* of that set and write it $\text{Ex}(S)$. Thus

$$\text{Ex}[G(x, y, z)] = 8(x^4 + y^4 + z^4 - 3y^2z^2 - 3z^2x^2 - 3x^2y^2). \tag{2.12}$$

This can take both positive and negative values according to the choice of $x$, $y$ and $z$. Clearly, $\sum_j \text{Ex}(S_j) = \text{Ex}(\sum_j S_j)$, where the notation $\sum_j S_j$ means that points which belong to more than one set $S_j$ contribute to the sum each time they occur. The notation thus does not denote the "union" of sets in the

usual sense. Furthermore, if a number of sets $S_1, S_2, \cdots, S_m$ (say) satisfy, either separately or together, the conditions for a second order rotatable arrangement except for the condition (2.5), then the condition

$$\text{Ex}(S_1 + \cdots + S_m) = \text{Ex}(S_1) + \cdots + \text{Ex}(S_m) = 0 \tag{2.13}$$

is a necessary and sufficient condition for the points of the whole set $S_1 + S_2 + \cdots + S_m$ to form a rotatable arrangement of second order. We shall make use of this important fact in Section 3.

For certain special choices of $(x, y, z)$ in three dimensions, the 24 points of $G(x, y, z)$ will coincide in pairs or in triplets or in quadruplets. For example, $G(p, q, 0)$ consists of the twelve points

$$(\pm p, \pm q, 0), \qquad (0, \pm p, \pm q), \qquad (\pm q, 0, \pm p), \tag{2.14}$$

each occurring twice. We may denote the 12 point set by $\frac{1}{2}G(p, q, 0)$. This set has excess

$$\text{Ex}[\tfrac{1}{2}G(p, q, 0)] = 4(p^4 + q^4 - 3p^2q^2), \tag{2.15}$$

a quantity which may be made positive or negative according to the values of $p$ and $q$.

The set $\frac{1}{2}G(p, q, 0)$ will itself form a rotatable arrangement if $p^4 - 3p^2q^2 + q^4 = 0$ or $p^2/q^2 = (3 \pm \sqrt{5})/2$. Thus $p/q = \theta$ or $\theta^{-1}$ where $\theta = (\sqrt{5} + 1)/2$, $\theta^{-1} = (\sqrt{5} - 1)/2$. Thus the set reduces to the 12 points $(\pm\theta, \pm 1, 0)$, $(\pm 1, 0, \pm\theta)$, $(0, \pm\theta, \pm 1)$, which as Coxeter [8] shows constitute the vertices of an icosahedron. Adding center points we get the icosahedron design given by Box and Hunter [3].

**3. The formation of rotatable arrangements and rotatable designs by combination of several generated points sets.** Consider the set $G(a, a, a)$; this consists of the eight points

$$(\pm a, \quad \pm a, \quad \pm a) \tag{3.1}$$

each occurring three times. We may therefore denote this set of 8 points by $\frac{1}{3}G(a, a, a)$.

$$\text{Ex}[\tfrac{1}{3}G(a, a, a)] = -16a^4, \tag{3.2}$$

which is always negative, hence this set alone cannot form a rotatable arrangement.

Consider the set $G(c, 0, 0)$; this consists of the six points

$$(\pm c, 0, 0), \qquad (0, \pm c, 0), \qquad (0, 0, \pm c) \tag{3.3}$$

each occurring four times. The six points may be denoted by $\frac{1}{4}G(c, 0, 0)$.

$$\text{Ex}[\tfrac{1}{4}G(c, 0, 0)] = 2c^4, \tag{3.4}$$

which is always positive, and so this set alone cannot form a rotatable arrangement.

For consistency of notation we may write the point (0,0,0) as $(1/24)G(0,0,0)$. Hence $n_0$ center points may be denoted by

$$\frac{n_0}{24}\, G(0, 0, 0). \tag{3.5}$$

Consider the combination of sets $\frac{1}{8}G(a, a, a)$ and $\frac{1}{4}G(c, 0, 0)$. Then

$$\text{Ex}[\tfrac{1}{8}G(a, a, a) + \tfrac{1}{4}G(c, 0, 0)] = -16a^4 + 2c^4. \tag{3.6}$$

This is zero if $c^2 = 2\sqrt{2}a^2$, in which case the 14 points form a rotatable arrangement. The actual design points are obtained by applying the scaling condition $\lambda_2 = 1$. This gives $8a^2 + 2c^2 = N = 14 + n_0$, where $n_0$ is the number of center points added. Thus $4(2 + \sqrt{2})a^2 = N$, and both $a$ and $c$ are determined when $N$ is given. We have obtained the well-known cube plus octahedron design first presented by Box and Hunter [3].

The method may now be extended. We have seen that the combination of generated sets leads to a single design when only two parameters are present, as in the example just given, since the two conditions $\text{Ex(set)} = 0, \lambda_2 = 1$, completely determine the design. The first condition alone completely determines the ratio of the two parameters and is sufficient to determine the design apart from scale. We now examine a combination of sets which contains three parameters. We shall see that we obtain a single infinity of designs which depend on a single parameter ratio. Consider the 20 points

$$\tfrac{1}{4}G(c_1, 0, 0), \qquad \tfrac{1}{4}G(c_2, 0, 0), \qquad \tfrac{1}{8}G(a, a, a).$$

The excess of this whole set is $2c_1^4 + 2c_2^4 - 16a^4$. Note that since $\text{Ex}[\frac{1}{8}G(a, a, a)] = -16a^4$ is negative, we must combine with it sets at least one of which has positive excess to compensate. Thus the set has zero excess if $c_1^4 + c_2^4 = 8a^4$. Set $c_1^2 = xa^2$, $c_2^2 = ya^2$. Then $x^2 + y^2 = 8$. Any positive values of $x$ and $y$ which satisfy this equation will give rise to a rotatable arrangement of second order. Thus if $(x, y)$ is a point of the circle $x^2 + y^2 = 8$ and also lies in the first quadrant, then we shall have a rotatable arrangement. No additional center points are required to make the arrangement into a design since three radii of the parts of the arrangement: $x^{1/2}a$, $y^{1/2}a$ and $\sqrt{3}a$ are not all equal. Now $N\lambda_2 = 2c_1^2 + 2c_2^2 + 8a^2 = 2(x + y + 4)a^2$. Applying the scaling condition $\lambda_2 = 1$, we obtain

$$y = \sqrt{8 - x^2} \qquad a = [N/2(x + y + 4)]^{1/2}, \qquad c_1 = x^{1/2}a, \qquad c_2 = y^{1/2}a,$$

and the design becomes completely determined. For this class, $\lambda_4/\lambda_2^2 = 8a^4/N$. We now derive, as special cases of the infinite class just obtained, two designs which were previously known.

(1) $x = 0$. Then $y = 2\sqrt{2}$, $c_2 = c^{\frac{1}{4}}a$, $c_1 = 0$. We have obtained the cube plus octahedron design, with 6 center points which are vertices of the degenerate octahedron.

(2) $x = y = 2$. Then $c_1 = c_2 = a\sqrt{2}$. This gives rise to the design described by Gardiner, Grandage and Hader, which consists of the vertices of a cube plus those of a doubled octahedron ([9], Sec. 6, Par. 6, first stage).

The first summary table which occurs in Section 5 contains several other infinite classes of this type.

**4. Classes of designs using sets with variable excess.** In the previous section the sets we used in combination had a positive or a negative excess. Let us now consider the set of 12 points $\frac{1}{2}G(p, q, 0)$. The excess of this set is $4(p^4 + q^4 - 3p^2q^2)$, a quantity which may be made positive or negative according to the way $p$ and $q$ are chosen. Thus $\frac{1}{2}G(p, q, 0)$ may be combined with all of the sets $\frac{1}{3}G(a, a, a)$, $\frac{1}{4}G(c, 0, 0)$ and $\frac{1}{2}G(f, f, 0)$ to obtain rotatable arrangements and hence designs. For example, $\mathrm{Ex}[\frac{1}{2}G(p, q, 0) + \frac{1}{3}G(a, a, a)] = 0$ if $p^4 + q^4 = 3p^2q^2 + 4a^4$. Set $p^2 = xa^2$, $q^2 = ya^2$, and we have $x^2 - 3xy + y^2 = 4$. Any point of this hyperbola which lies in the first quadrant will give rise to a rotatable arrangement of second order. If we solve for $y$ in terms of $x$, we obtain $y = [3x \pm \sqrt{5x^2 + 16}]/2$. This yields two positive solutions if $x > 2$; otherwise only one positive solution arises. This may easily be seen from Fig. 2. The radii of the separate point sets are $\sqrt{x + y}a$ and $\sqrt{3}a$ and these are equal when $x + y = 3$. Since the straight line $x + y = 3$ intersects the hyperbola in two points $P$ and $Q$, equality in (1.3) occurs for two arrangements of the class. For these two arrangements, the addition of center points is necessary to satisfy the non-singularity condition. Applying the scaling condition $\lambda_2 = 1$, we obtain an infinite class of second order rotatable designs, each design consisting of 20 points plus any center points which may have been added. The class depends on one parameter $x$. Given $x \geqq 0$,

$$y = [3x \pm \sqrt{5x^2 + 16}]/2, \quad a = [N/4(x + y + 2)]^{1/2}, \quad p = x^{1/2}a, \quad q = y^{1/2}a,$$

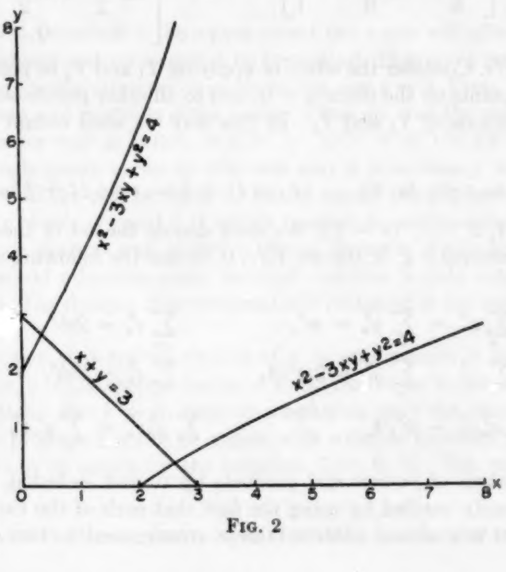


FIG. 2

where the lower sign in $y$ is to be taken only when $x > 2$. For this class

$$\lambda_4/\lambda_2^2 = (8a^4 + 4p^2q^2)/N = 4(2 + xy)a^4/N.$$

This class has two well-known special cases.

(1) When $a = 0$, $x = \infty$, $y = \infty$. Ignoring the degenerate set $\frac{1}{2}G(a, a, a)$, we obtain the icosahedron design discussed at the end of Section 2.

(2) If we choose one of the two points on the hyperbola for which $x + y = 3$, then $x = y^{-1} = (3 \pm \sqrt{5})/2 = \theta^2, \theta^{-2}$, where $\theta = (\sqrt{5} + 1)/2$. Thus the 20 design points (other than the center points) consist of constant multiples of

$$(0, \pm\theta^{-1}, \pm\theta), \qquad (\pm\theta, 0, \pm\theta^{-1}), \qquad (\pm\theta^{-1}, \pm\theta, 0), \qquad (\pm 1, \pm 1, \pm 1).$$

As Coxeter [8] shows, these are the vertices of a dodecahedron, which form a well-known second order rotatable design, given in [3].

Several other classes of this type may be found in the summary table.

**5. Summary table.** Table I is a table of infinite classes of second order rotatable designs in three dimensions of the type derived in Sections 3 and 4. The table shows the generated sets used to form each class together with the design coordinate values in terms of a single parameter.

**6. A second method of generating point sets suitable for building second order rotatable designs.** Define

$$T_1 = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad T_2 = \begin{bmatrix} \cos\frac{\alpha}{2} & \sin\frac{\alpha}{2} & 0 \\ \sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

where $\alpha = 2\pi/s$. Consider the effect of applying $T_1$ and $T_2$ to points of the form $(r, 0, b)$, i.e., points on the plane $y = 0$, and to all other points obtained from repeated applications of $T_1$ and $T_2$. In this way we shall obtain $2s$ points with coordinates

(6.1) $\quad (r\cos t\alpha, r\sin t\alpha, b), \qquad (r\cos(t + \frac{1}{2})\alpha, r\sin(t + \frac{1}{2})\alpha, -b),$

where $t = 0, 1, 2, \cdots, (s - 1)$. We shall denote the set of these $2s$ points by $T_s(r, 0, b)$. Provided $s \geq 5$, the set $T_s(r, 0, b)$ has the following sums of powers and products:

$$\sum_u x_u^2 = \sum_u y_u^2 = sr^2, \qquad \sum_u z_u^2 = 2sb^2,$$
$$\sum_u x_u^4 = \sum_u y_u^4 = 3sr^4/4, \qquad \sum_u z_u^4 = 2sb^4,$$
$$\sum_u x_u^2y_u^2 = sr^4/4, \qquad \sum_u y_u^2z_u^2 = \sum_u z_u^2x_u^2 = sr^2b^2,$$

and all other sums of powers and products up to and including order four are zero. This is easily verified by using the fact that each of the two $s$-gons in the set of $2s$ points is a second order rotatable arrangement in two dimensions [3].

A rotation about the $z$ axis of the complete point arrangement will not affect the properties held by the sums of powers and products. We now recall the cyclic group $W$, $W^2$, $I$, defined in Section 2, and apply this to $T_s(r, 0, b)$ to give set $T_x(b, r, 0)$ and $T_y(0, b, r)$. In all we now have $6s$ points, which we denote by $T(r, 0, b)$ with coordinates

$$\begin{array}{ll} (r \cos t\alpha, r \sin t\alpha, b), & (r \cos (t + \tfrac{1}{2})\alpha, r \sin (t + \tfrac{1}{2})\alpha, -b), \\ (b, r \cos t\alpha, r \sin t\alpha), & (-b, r \cos (t + \tfrac{1}{2})\alpha, r \sin (t + \tfrac{1}{2})\alpha), \\ (r \sin t\alpha, b, r \cos t\alpha), & (r \sin (t + \tfrac{1}{2})\alpha, -b, r \cos (t + \tfrac{1}{2})\alpha), \end{array}$$

where $\alpha = 2\pi/s$ $(s \geqq 5)$ and $t = 0, 1, 2, \cdots, (s - 1)$.

The set $T(r, 0, b)$ has sums of powers and products

$$\sum_u x_u^2 = \sum_u y_u^2 = \sum_u z_u^2 = 2s(r^2 + b^2),$$

$$\sum_u x_u^4 = \sum_u y_u^4 = \sum_u z_u^4 = s(3r^4 + 4b^4)/2,$$

$$\sum_u x_u^2 y_u^2 = \sum_u y_u^2 z_u^2 = \sum_u z_u^2 x_u^2 = sr^2(r^2 + 8b^2)/4,$$

and all other sums of powers and products up to and including order four are zero.

The formulae for the sums of powers and products will extend to the case $s = 4$, provided we fix as the set $T_s(r, 0, b)$ the points

$$(\pm r, 0, b), \qquad (0, \pm r, b), \qquad (\pm r/\sqrt{2}, \pm r/\sqrt{2}, -b).$$

In the case $s = 4$, rotation of the $s$-gons about the $z$ axis will affect the sums of powers and products and thus cannot be permitted. This point must be remembered whenever specific reference is made to the case $s = 4$. From the properties of sums of powers and products given above, it follows that the excess of the set, defined in the same way as before, is $s(3r^4 - 24r^2b^2 + 8b^4)/4$. Of course the excess of each single point varies in this case and it is necessary to consider the total effect over all the points. Since its excess can be made positive or negative according to the choice of $r$ and $b$, it will be possible to combine the set $T(r, 0, b)$ with sets of both positive and negative excess. Because of the large number of points which would otherwise arise, we shall combine it only with $\frac{1}{2}G(a, a, a)$ and $\frac{1}{4}G(c, 0, 0)$. The designs thus obtained will be found in the second summary table below.

In the same way that special choices of $x$, $y$, and $z$ made it possible to take fractions of $G(x, y, z)$, a special choice of $b$ will enable us to use a smaller point set than $T(r, 0, b)$. Set $b = 0$; then by employing only the transformation $T_1$ and $W$ we can produce a set of $3s$ points with suitable moment properties. We shall denote these $3s$ points by the notation $T_0(r, 0, 0)$. The points will have coordinates

$$(r \cos t\alpha, r \sin t\alpha, 0), \qquad (r \sin t\alpha, 0, r \cos t\alpha), \qquad (0, r \cos t\alpha, r \sin t\alpha),$$

## TABLE I

| Set ...... | $G(p, q, r)$ | $\frac{1}{2}G(p, q, 0)$ | $\frac{1}{3}G(a,a,a)$ | $\frac{1}{4}G(c,0,0)$ | Number of points in design class | Range of first parameter ratio on which class depends | Second parameter ratio in terms of first | Design point coordinate values in terms of $N$ and parameter ratios | Value of $\lambda_4/\lambda_2^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Points.... | $(\pm p, \pm q, \pm r)$ etc. | $(\pm p, \pm q, 0)$ etc. | $(\pm a, \pm a, \pm a)$ | $(\pm c, 0, 0)$ etc. | | | | | |
| No. of points.. | 24 | 12 | 8 | 6 | | | | | |
| | Number of sets used for class | | | | | | | | |
| | 1 | | | | 24 | $0 \leqq t \leqq (3 - \sqrt{5})/2$ <br> $t \geqq (3 + \sqrt{5})/2$ | $s = \frac{1}{2}[3(t + 1) \pm \sqrt{5(t^2 + 6t + 1)}]$ | $r = [N/8(s + t + 1)]^{1/2}$ <br> $p = s^{1/2}r, \quad q = t^{1/2}r$ | $8(st + s + t)r^4/N$ |
| | | | 1 | 2 | 20 | $0 \leqq x \leqq 2\sqrt{2}$ | $y = \sqrt{8 - x^2}$ | $a = [N/2(x + y + 4)]^{1/2}$ <br> $c_1 = x^{1/2}a, \quad c_2 = y^{1/2}a$ | $8a^4/N$ |
| | | | 2 | 1 | 22 | $0 \leqq x \leqq \frac{1}{2\sqrt{2}}$ | $y = \sqrt{\frac{1}{4} - x^2}$ | $c = [N/2(4x + 4y + 1)]^{1/2}$ <br> $a_1 = x^{1/2}c, \quad a_2 = y^{1/2}c$ | $c^4/N$ |
| | | 1 <br> $p = q = f$ | | 2 | 24 | $0 \leqq x \leqq 1$ | $y = \sqrt{2 - x^2}$ | $f = [N/2(x + y + 4)]^{1/2}$ <br> $c_1 = x^{1/2}f, c_2 = y^{1/2}f$ | $4f^4/N$ |
| | | 1 <br> $p = q = f$ | 1 | 1 | 26 | $0 \leqq x \leqq \frac{1}{2\sqrt{2}}$ | $y = [(1 - 2x^2)/8]^{1/2}$ | $c = [N/2(4x + 4y + 1)]^{1/2}$ <br> $f = x^{1/2}c, a = y^{1/2}c$ | $4(x^2 + 2y^2)c^4/N$ |
| | | 1 | 1 | | 20 | $x \geqq 0$ | $y = \frac{1}{2}[3x \pm \sqrt{5x^2 + 16}]$ | $a = [N/4(x + y + 2)]^{1/2}$ <br> $p = x^{1/2}a, q = y^{1/2}a$ | $4(2 + xy)a^4/N$ |
| | | 1 | | 1 | 18 | $x \geqq 0.63$ | $y = \frac{1}{2}[3x \pm \sqrt{5x^2 - 2}]$ | $c = [N/2(2x + 2y + 1)]^{1/2}$ <br> $p = x^{1/2}c, q = y^{1/2}c$ | $4xyc^4/N$ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 2 (one $p = q = f$) | | | 24 | $x \geq 0$ | $y = \frac{1}{2}[3x \pm \sqrt{5x^2 + 4}]$ | $f = [N/4(x + y + 2)]^{1/2}$ $p = x^{1/2}f,\ q = y^{1/2}f$ | $4(xy + 1)f^4/N$ |
| | 1 $q = r$ | | 1 | | 32 | $x \geq \sqrt{2}$ | $y = \sqrt{10x^2 - 2} - 3x$ | $a = [N/8(x + 2y + 1)]^{1/2}$ $p = x^{1/2}a,\ q = y^{1/2}a$ | $8(2xy + y^2 + 1)a^4/N$ |
| | 1 $q = r$ | | | 1 | 30 | $x \geq 0$ | $y = \sqrt{10x^2 + \frac{1}{4}} - 3x$ | $c = [N/2(4x + 8y + 1)]^{1/2}$ $p = x^{1/2}c,\ q = y^{1/2}c$ | $2(8xy + 4y^2 + 1)c^4/N$ |

where $t = 0, 1, 2, \cdots, (s - 1)$ and $s \geqq 5$. The sums of powers and products of the set are

$$\sum_u x_u^2 = \sum_u y_u^2 = \sum_u z_u^2 = sr^2,$$
$$\sum_u x_u^4 = \sum_u y_u^4 = \sum_u z_u^4 = 3sr^4/4,$$
$$\sum_u x_u^2 y_u^2 = \sum_u y_u^2 z_u^2 = \sum_u z_u^2 x_u^2 = sr^4/8,$$

and all other sums of powers and products up to and including order four are zero.

Clearly any rotation of the 3 $s$-gons about their axes will also give rise to the same moments, but we shall restrict attention here to the set $T_0(r, 0, 0)$. From the sums of powers and products it follows that the excess of this set is $3sr^4/8$ which is a positive excess. Thus to form an infinite class of second order designs we must combine $T_0(r, 0, 0)$ with sets at least one of which has negative excess Two examples of this will be found in Table II.

**7. An extension of the method: a 16 point design class.** Consider the set of 12 points

$$\begin{array}{llll} (x, y, z), & (x, -y, -z), & (-x, y, -z), & (-x, -y, z), \\ (y, z, x), & (-y, -z, x), & (y, -z, -x), & (-y, z, -x), \\ (z, x, y), & (-z, x, -y), & (-z, -x, y), & (z, -x, -y). \end{array} \tag{7.1}$$

This set consists of all points of $G(x, y, z)$ for which the product of the coordinates is $xyz$. It can be described as a $\frac{1}{2}$ replicate of $G(x, y, z)$ and we shall write it

$$G^{(+\frac{1}{2})}(x, y, z). \tag{7.2}$$

The complementary set, where the product of the coordinates is $-xyz$, we shall denote by

$$G^{(-\frac{1}{2})}(x, y, z). \tag{7.3}$$

The set (7.2) satisfies all the conditions for a second order rotatable arrangement except two. These are

$$\sum_{u=1}^{N} x_{iu}^4 = 3\sum_{u=1}^{N} x_{iu}^2 x_{ju}^2, \qquad (i, j = 1, 2, 3), \qquad (i \neq j) \tag{7.4}$$

and

$$\sum_{u=1}^{N} x_{1u}x_{2u}x_{3u} = 0. \tag{7.5}$$

We recall that

$$\text{Ex[Point set}(x_{1u}, x_{2u}, x_{3u}), \quad u = 1, 2, \cdots, N] = \sum_{u=1}^{N} x_{iu}^4 - \sum_{u=1}^{N} x_{iu}^2 x_{ju}^2, \qquad (i, j = 1, 2, 3), \qquad (i \neq j) \tag{7.6}$$

## TABLE II

| Set...... | $T(r,0,b)$ | $T_s(r,0,0)$ | $\frac{1}{2}G(a,a,a)$ | $\frac{1}{2}G(0,0,c)$ | Number of points in design class | Range of first parameter ratio on which class depends | Second parameter ratio in terms of first | Design point coordinate values in terms of N and parameter ratios | Value of $\lambda_4/\lambda_2^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Points.... | ($r\cos t\alpha$, $r\sin t\alpha$, $b$ etc.) | ($r\cos t\alpha$, $r\sin t\alpha$, 0) etc. | ($\pm a,\pm a,\pm a$) | ($\pm c$,0,0) (0,$\pm c$,0) (0,0,$\pm c$) | | | | | |
| No. of points.. | $s \geqq 4$ $6s$ | $s \geqq 5$ $3s$ | 8 | 6 | | | | | |
| | Number of sets used for class | | | | | | | | |
| | 1 | | 1 | | $6s + 8$ $s \geqq 4$ | $x \geqq 0$ | $y = \frac{1}{4}[6x \pm \sqrt{30x^2 + 128/s}$ | $a = [N/2(s(x+y)+4)]^{1/2}$ $r = x^{1/2}a,\ b = y^{1/2}a$ | $\dfrac{[sx(x+8y)+32]a^4}{4N}$ |
| | 1 | | | 1 | $6s + 6$ $s \geqq 4$ | $x \geqq \sqrt{8/15s}$ | $y = \frac{1}{4}[6x \pm \sqrt{30x^2 - 16/s}$ | $c = [N/2(s(x+y)+1)]^{1/2}$ $r = x^{1/2}c,\ b = y^{1/2}c$ | $\dfrac{[sx(x+8y)+8]c^4}{4N}$ |
| | | 1 | 2 | | $3s + 16$ $s \geqq 5$ | $0 \leqq x \leqq \sqrt{3s/256}$ | $y = \sqrt{3s/128 - x^2}$ | $r = [N/(8x + 8y + s)]^{1/2}$ $a_1 = x^{1/2}r,\ a_2 = y^{1/2}r$ | $5sr^4/16N$ |
| | | 1 | 1 | 1 | $3s + 14$ $s \geqq 5$ | $0 \leqq x \leqq \sqrt{128/3s}$ | $y = \frac{1}{4}\sqrt{128 - 3sx^2}$ | $a = [N/(sx + 2y + 8)]^{1/2}$ $r = x^{1/2}a,\ c = y^{1/2}a$ | $(64 + sx^2)a^4/8N$ |

Let us define a second excess function which relates to the left member of (7.5) as

(7.7) $$\mathrm{Fx}[\text{Point set } (x_{1u}, x_{2u}, x_{3u}), \quad u = 1, 2, \cdots, N] = \sum_{u=1}^{N} x_{1u}x_{2u}x_{3u}.$$

Then if $S$ is a point set or a combination of points sets which satisfies all of conditions (1.2) except (7.4) and (7.5), and if

(7.8) $$\mathrm{Ex}(S) = 0, \qquad Fx(S) = 0,$$

then $S$ is a rotatable arrangement of the second order. Now

(7.9) $$\mathrm{Ex}[G^{(\pm\frac{1}{2})}(x, y, z)] = 4(x^4 + y^4 + z^4 - 3y^2z^2 - 3z^2x^2 - 3x^2y^2)$$

(7.10) $$\mathrm{Fx}[G^{(\pm\frac{1}{2})}(x, y, z)] = \pm 12\, xyz.$$

The set $G^{(+\frac{1}{2})}(a, a, a)$ consists of the four points

(7.11) $$(a, a, a), \quad (a, -a, -a), \quad (-a, a - a), \quad (-a, -a, a),$$

each repeated three times. Thus we may denote the four points (7.11) which form a half replicate of the $2^3$ factorial design, by $\frac{1}{3}G^{(\frac{1}{2})}(a, a, a)$. Similarly the set $\frac{1}{3}G^{(-\frac{1}{2})}(a, a, a)$ consists of the four point

$$(-a, -a, -a), \qquad (-a, a, a), \qquad (a, -a, a), \qquad (a, a, -a).$$

It is easily seen that

$$\mathrm{Ex}[\tfrac{1}{3}G^{(\pm\frac{1}{2})}(a, a, a)] = -8a^4, \qquad \mathrm{Fx}[\tfrac{1}{3}G^{(\pm\frac{1}{2})}(a, a, a)] = \pm 4a^3.$$

Let $S$ be the set of 16 points defined by

$$S = G^{(+\frac{1}{2})}(x, y, z) + \tfrac{1}{3}G^{(-\frac{1}{2})}(a, a, a),$$

$$\mathrm{Ex}(S) = -8a^4 + 4(x^4 + y^4 + z^4 - 3y^2z^2 - 3z^2x^2 - 3x^2y^2),$$

$$\mathrm{Fx}(S) = 12\, xyz - 4a^3.$$

Thus $S$ is a rotatable arrangement if

(7.12) $$x^4 + y^4 + z^4 - 3(y^2z^2 + z^2x^2 + x^2y^2) = 2a^4, \qquad 3\, xyz = a^3.$$

If we set

(7.13) $$x^2 = ua^2, \qquad y^2 = va^2, \qquad z^2 = wa^2,$$

it follows from (7.12) that we can write

$$u + v + w = \beta, \qquad uv + vw + wu = (\beta^2 - 2)/5, \qquad uvw = 1/9.$$

These equations imply that $u$, $v$ and $w$ are the roots of the cubic

(7.14) $$t^3 - \beta t^2 + (\beta^2 - 2)t/5 - 1/9 = 0.$$

If for a given $\beta$ this cubic has three positive roots $u$, $v$ and $w$, we shall be able to

use these values to obtain a rotatable arrangement of the second order which contains only 16 points, using the relations (7.13). A sufficient condition for

$$Ax^3 + Bx^2 + Cx + D = 0 \tag{7.15}$$

to have three positive roots (provided all roots are real) is $A > 0$, $B < 0$, $C > 0$, $D < 0$. Thus if $\beta > \sqrt{2}$ and all three roots of (7.14) are real, they are all positive. The necessary and sufficient condition for (7.15) to have three real roots is $\Delta = B^2C^2 + 18ABCD - 4AC^3 - 27A^2D^2 - 4B^3D > 0$ (see Conkwright [7]). For the equation (7.14) we find

$$\Delta(\beta) = 3645(9\beta^6 + 36\beta^4 - 50\beta^3 - 252\beta^2 - 900\beta - 87). \tag{7.16}$$

It may be shown that $\Delta(2.691376)/3645 = .0031$, $\Delta(2.691375)/3645 = -.04$, so that a root of $\Delta = 0$ lies near $\beta = 2.691376$. Furthermore

$$\Delta(2.691376 + s)/3645 = .0031 + \Delta_1(s),$$

where $\Delta_1(s)$ is the following sixth degree polynomial in $s$ with all coefficients positive:

$$\Delta_1(s) = 9s^6 + 145.3s^5 + 1013.9s^4 + 3846.7s^3 + 7992.1s^2 + 7089.8s$$

Hence $s > 0 \Rightarrow \Delta_1(s) > 0 \Rightarrow \Delta(2.691376 + s) > 0$, and

$$\Delta(\sqrt{2})/3645 = -1789, \; \Delta''(\beta)/7290 = 135\beta^4 + 216\beta^2 - 150\beta - 250 > 0$$

for $\beta > \sqrt{2}$.

TABLE III

*A Selection of Designs from the 16 Point Series (when $n_0 = 0$)*

| $\beta$ | $a$ | $x$ | $y$ | $z$ | $\lambda_4/\lambda_2^2$ |
|---|---|---|---|---|---|
| 2.691376 | 1.04096 | .49090 | .49090 | 1.56026 | .60140 |
| 2.7 | 1.03975 | .45968 | .52238 | 1.56036 | .60131 |
| 3 | 1.00000 | .31645 | .67348 | 1.56405 | .60000 |
| 4 | .89443 | .18375 | .82366 | 1.57775 | .60800 |
| 5 | .81650 | .12862 | .88669 | 1.59078 | .62222 |
| 6 | .75593 | .09737 | .92330 | 1.60206 | .63673 |
| 7 | .70711 | .07722 | .94697 | 1.61160 | .65000 |
| 8 | .66667 | .06328 | .96348 | 1.61965 | .66173 |
| 9 | .63246 | .05321 | .97559 | 1.62647 | .67200 |
| 11 | .57735 | .03951 | .99212 | 1.63732 | .68889 |
| 14 | .51640 | .02767 | 1.00687 | 1.64887 | .70756 |
| 19 | .44721 | .01759 | 1.02001 | 1.66110 | .72800 |
| 49 | .28284 | .00430 | 1.04018 | 1.68464 | .76928 |
| 99 | .20000 | .00151 | 1.04601 | 1.69288 | .78432 |
| $\infty$ | 0 | 0 | 1.05146 | 1.70130 | .80000 |

When $n_0 = 0$, multiply $a$, $x$, $y$ and $z$ by $\alpha$ and multiply $\lambda_4/\lambda_2^2$ by $\alpha^2$, where $\alpha^2 = 1 + (n_0/16)$.

The variation in the values of $a$ $x$, $y$ and $z$ is so well controlled that it is possible to use a graph to find their values for values of $\beta$ other than those in the table.

This means that the function $\Delta$ is convex for $\beta > \sqrt{2}$ and thus has only one root in that range which must be at approximately $\beta = 2.691376$. Thus if $\beta > 2.7$ the equation (7.14) gives rise to three real positive roots $u$, $v$ and $w$ and the 16 points of $S$ form a second order rotatable arrangement. The radii of the two sets of points which comprise the arrangement are $\sqrt{\beta}\, a$ and $\sqrt{3}\, a$. Thus, when $\beta = 3$ it will be necessary to add center points to the arrangement in order to satisfy the non-singularity condition. It is desirable to add center points to arrangements which arise from values of $\beta$ near the singular value 3 in order that the variances of the estimates of the model coefficients will not be large. When $a = 0$, we shall retain the degenerate points as center points. If $N = 16 + n_0$ where $n_0$ is the number of center points added, it is easy to verify that the scaling condition $\lambda_2 = 1$ leads to $a^2 = N/4(\beta + 1)$. Thus we have found an infinite class of second order rotatable designs depending on a parameter $\beta$; each design contains 16 points excluding any center points which may have been added. Given a value of $\beta > 2.691376$, we can find $u$, $v$ and $w$, the positive roots of (7.14). Then

$$a = [N/4(\beta + 1)]^{\frac{1}{2}}, \qquad x = u^{\frac{1}{2}}a, \qquad y = v^{\frac{1}{2}}a, \qquad z = w^{\frac{1}{2}}a,$$

and the design is completely determined. An easy calculation shows that

$$\lambda_4/\lambda_2^2 = (\beta^2 + 3)N/20(\beta + 1)^2. \tag{7.17}$$

Table III contains some of the designs of this series. The table was obtained by substituting for $\beta$ in (7.14) a specific value and solving the cubic equation. Only the range $\beta > 2.691376$ need be considered. The values given for $x$, $y$, $z$ and $a$ are those to be used when $n_0 = 0$, i.e., when no center points are added; for $n_0$ center points these values must be multiplied by the factor $\alpha = [1 + (n_0/16)]^{\frac{1}{2}}$. The design points are obtained from (7.1) and (7.11) with appropriate values for $x$, $y$, $z$ and $a$ from the table. The value of $\lambda_4/\lambda_2^2$ in the table is calculated from (7.17) when $N = 16$. For $n_0$ center points these values must be multiplied by $\alpha^2 = 1 + (n_0/16)$.

## REFERENCES


[1] G. E. P. Box, "Multi-factor designs of first order," *Biometrika*, Vol. 39 (1952), pp. 49–57.

[2] G. E. P. Box, "The exploration and exploitation of response surfaces: Some general considerations and examples," *Biometrics*, Vol. 10 (1954), pp. 16–60.

[3] G. E. P. Box and J. S. Hunter, "Multi-factor experimental designs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 195–241.

[4] G. E. P. Box and J. S. Hunter, "Experimental designs for the exploration and exploitation of response surfaces," *Proceedings of Symposium on Design of Industrial Experiments*, Institute of Statistics, Raleigh, North Carolina, 1956.

[5] G. E. P. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," *J. Roy. Stat. Soc.*, (*Ser.* B), Vol. 13 (1951), pp. 1–45.

[6] Richard L. Carter, "New designs for the exploration of response surfaces," Ph.D. Thesis, University of North Carolina, Chapel Hill, North Carolina, 1957.

[7] N. B. Conkwright, *Introduction to the Theory of Equations*, Ginn, Boston, 1941.

[8] H. S. M. Coxeter, *Regular Polytopes*, Methuen and Co., Ltd., London, 1948.

[9] D. A. Gardiner, A. H. E. Grandage and R. J. Hader, "Third order rotatable designs for exploring response surfaces," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1082–1096.

# THE PROBABILITY IN THE EXTREME TAIL OF A CONVOLUTION[1]

David Blackwell and J. L. Hodges, Jr.

*University of California, Berkeley*

**1. Summary.** Let $X_1, X_2, \cdots$ be independent and identically distributed random variables with possible values that are integers whose differences have g.c.d. one. Assume the m.g.f. of $X_1$ exists in an interval about 0, let $a$ be any number such that $E(X_1) < a < \sup X_1$, and let $\phi(a, t) = Ee^{t(X_1 - a)}$. There exists a unique value $t^*(a)$ of $t$ which minimizes $\phi(a, t)$ with respect to $t$; write $m(a) = \phi[a, t^*(a)]$ and $z = e^{-t^*(a)}$. Let $Y_1, Y_2, \cdots$ be independent and identically distributed random variables such that $Y_1$ and $X_1$ have the same range and $\Pr(Y_1 = x) = \Pr(X_1 = x) \cdot e^{t^*(a)(x-a)}/m(a)$, and let $\mu_2 = \sigma^2, \mu_3, \mu_4$ be central moments of $Y_1$.

We show that $\Pr\{X_1 + \cdots + X_n = na\} = [m(a)]^n \Pr\{Y_1 + \cdots + Y_n = na\}$, and use this to establish the approximation $\Pr\{X_1 + \cdots + X_n = na\} = \pi_n^{**}[1 + 0(n^{-2})]$, where $na$ is a possible value of $X_1 + \cdots + X_n$ and

$$\pi_n^{**} = \frac{[m(a)]^n}{\sigma\sqrt{2\pi n}}\left[1 + \frac{1}{8n}\left(\frac{\mu_4}{\mu_2^2} - 3 - \frac{5}{3}\frac{\mu_3^2}{\mu_2^3}\right)\right].$$

Similarly we find that $\Pr\{X_1 + \cdots + X_n \geq na\} = \Pi_n^{**}[1 + 0(n^{-2})]$, where

$$\Pi_n^{**} = \pi_n^{**} \cdot \frac{1}{1-z}\left\{1 - \frac{1}{2n}\left[\frac{(z\mu_3/\mu_2) + z(1+z)/(1-z)}{(1-z)\mu_2}\right]\right\}.$$

We provide some numerical illustrations of the accuracy of these approximations, and give a conjectured analog of the leading term of $\Pi_n^{**}$ for nonlattice variables.

**2. Introduction.** Let $X_1, X_2, \cdots$ be independent identically distributed random variables whose common moment generating function $Ee^{tX_1}$ is finite in some interval about 0, and let $a$ be any number such that $E(X_1) < a < \sup X_1$. We shall be interested in the tail probability

$$\Pi_n(a) = \Pr\{X_1 + \cdots + X_n \geq na\}.$$

As $n \to \infty$ we shall of course have $\Pi_n(a) \to 0$, since $na$ exceeds the expected value of the sum by about $\sqrt{n}$ standard deviations. The study of the speed with which $\Pi_n(a) \to 0$ was initiated by Cramér [2] in 1938; his results were extended by Feller [3] and Chernoff [1].

Denote by $\phi(a, t)$ the moment generating function of $X_1 - a$: $\phi(a, t) = Ee^{t(X_1-a)}$. Chernoff shows that for each $a$ there is a unique value of $t$, say $t^*(a)$,

Received September 26, 1958; revised August 3, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53) and (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

for which $\phi$ achieves its minimum, and writes $\phi[a, t^*(a)] = m(a)$. He shows that for every $\epsilon > 0$,

$$[m(a) - \epsilon]^n \leqq \Pi_n(a) \leqq [m(a)]^n$$

for sufficiently large $n$, with the right inequality holding for all $n$.

This result establishes in a sense the speed with which $\Pi_n(a) \to 0$, but it is not precise enough to permit the approximation of $\Pi_n(a)$ with a small relative error, since the ratio of upper to lower bound tends to infinity with $n$. There remains the problem of developing a relatively accurate approximation for $\Pi_n(a)$. Cramér [2] has found such an approximation for the case in which $X_1$ has an absolutely continuous component. We are interested in the case of lattice variables, i.e., the case in which there are constants $A \neq 0$ and $B$ such that $AX_1 + B$ has only integer values.

**3. An identity.** In this section we restrict attention to sequences $\{X_n\}$ of discrete variables.

THEOREM 1: *Let $X_1, X_2, \cdots$ be independent identically distributed discrete variables whose common moment generating function $E(e^{tX_1})$ is finite for some interval about* 0. *For any $a$ with $E(X_1) < a < \sup X_1$, let*

$$m(a) = \min_t Ee^{t(X_1 - a)} = \min_t \phi(t, a) = \phi[t^*(a), a], \textit{ say,}$$

*and let $Y_1, Y_2, \cdots$ be independent identically distributed discrete variables whose common distribution is defined by*

$$\Pr\{Y_1 = x\} = \Pr\{X_1 = x\} \exp[t^*(a)(x - a)]/m(a) \quad \textit{for all } x.$$

*Then for all $n$,*

$$\Pr\{X_1 + \cdots + X_n = na\} = [m(a)]^n \Pr\{Y_1 + \cdots + Y_n = na\}.$$

The shift from the random variable $X$ to the random variable $Y$ has the effect of moving our event from the extreme tail to the center, since $na$ is just the expected value of $Y_1 + \cdots + Y_n$. This shift is not new. It is essentially carried out in Cramér's original paper. Wald [6] made a similar change in his "conjugate" distribution, introduced in the study of a problem arising in sequential analysis. Shannon [4] encountered the shift in a problem of information theory, and remarked (p. 15): "These tilted probabilities are convenient in evaluating the 'tails' of distribution that are sums of other distributions."

PROOF OF THEOREM 1: As noted by Chernoff, $\phi(t, a)$ is for each $a$ a strictly convex function of $t$ and attains its minimum at a unique $t = t^*(a)$. Write $p(x) = \Pr\{X_1 = x\}$. We have $\phi(a, t) = \sum_x p(x)e^{t(x-a)}$, so that

$$\phi_2[a, t^*(a)] = \sum_x (x - a)p(x)e^{t^*(a)(x-a)} = 0, \tag{1}$$

where $\phi_i$ denotes the partial derivative of $\phi$ with respect to its $i$th argument. Write $q(x) = p(x)e^{t^*(a)(x-a)}/m(a)$. Then $q(x)$ is a discrete probability distribution, and (1) asserts that the mean of the $q$ distribution is $a$. Let $Y_1, Y_2, \cdots$

be a sequence of independent identically distributed variables with common distribution $q$, and let $x_1, \cdots, x_n$ be any sequence of numbers whose sum is $na$. Then

$$\begin{aligned}
\Pr\{(Y_1, \cdots, Y_n) = (x_1, \cdots, x_n)\} &= q(x_1) \cdots q(x_n) \\
&= p(x_1) \cdots p(x_n) \exp [t^*(a)(x_1 + \cdots + x_n - na)]/[m(a)]^n \\
&= \Pr\{(X_1, \cdots, X_n) = (x_1, \cdots, x_n)\}/[m(a)]^n.
\end{aligned}$$

Summing over all sequences $(x_1, \cdots, x_n)$ such that $x_1 + \cdots + x_n = na$ yields the assertion of the theorem.

Theorem 1 extends to $M$-dimensional variables without change. We shall not use this extension, but give it for completeness.

THEOREM 2: *Let $X_1, X_2, \cdots$ be independent identically distributed discrete $M$-dimensional variables, and let $a$ be any interior point of the convex hull of the range of $X_1$, $a \neq \mu$, where $\mu = E(X_1)$. Suppose that there is a positive number $b$ such that the moment generating function $Ee^{t \cdot X_1}$ is finite for all $t$ for which $|t| \leqq b$ and $t \cdot (a - \mu) \geqq 0$ where, for any $t = (t_1, \cdots, t_m)$, $x = (x_1, \cdots, x_m)$, $t \cdot x$ denotes the inner product $\sum t_i x_i$. Then the moment generating function of $X_1 - a$ achieves its minimum value $m(a)$, say, at a unique $t = t^*(a)$, say, and, if $Y_1, Y_2, \cdots$ are independent identically distributed discrete $M$-dimensional variables whose common distribution is defined by $\Pr\{Y_1 = x\} = \Pr\{X_1 = x\} e^{t^* \cdot (x - a)}/m(a)$, then, $Y_1$ has mean $a$ and, for all $n$,*

$$\Pr\{X_r + \cdots + X_n = na\} = [m(a)]^n \Pr\{Y_1 + \cdots + Y_n = na\}.$$

The proof parallels that of Theorem 1. Again, the moment generating function has a minimum $m(a)$ at a unique $t^*$, at which $\partial\phi \mid \partial t_i = 0$ for all $i$. These equations assert that the $q$ distribution defined by $q(x) = \Pr(X_1 = x)e^{t^* \cdot (x-a)}/m(a)$ has mean $a$, and the rest of the proof is as before.

**4. The individual term.** In this section we shall specialize to the case of lattice variables. This means that it is possible by a linear transformation to assure that the values of $X_1$ are integers whose differences have g.c.d. 1; we assume this reduction has been carried out. We are then able to develop expressions for $\pi_n(a)$, using a method exploited for example by von Mises [5 Sec. 8].

Let $\sigma^2 = \mu_2, \mu_3, \mu_4$ be central moments of $Y$ of order 2, 3, 4. We shall establish

THEOREM 3: *If $X_1, X_2, \cdots$ are integer-valued variables satisfying the hypotheses of Theorem 1, the approximation*

$$\pi_n^*(a) = \frac{[m(a)]^n}{\sqrt{2\pi n}\sigma}$$

*for $\pi_n(a) = \Pr\{X_1 + \cdots + X_n = na\}$ has relative error of order $n^{-1}$, while the approximation*

$$\pi_n^{**}(a) = \pi_n^*(a)\left\{1 + \frac{1}{8n}\left[\frac{\mu_4}{\mu_2^2} - 3 - \frac{5}{3}\frac{\mu_3^2}{\mu_2^3}\right]\right\}$$

*for $\pi_n(a)$ has relative error of order $n^{-2}$.*

PROOF: In general, if a random variable $U$ with characteristic function $\eta$ has only integral values it is easy to check [5] that

$$\Pr(U = u) = (1/2\pi) \int_{-\pi}^{\pi} e^{-itu} \eta(t) \, dt.$$

Since $Y_1 + \cdots + Y_n$ is such a random variable, we have

$$\Pr(Y_1 + \cdots + Y_n = na) = (1/2\pi) \int_{-\pi}^{\pi} e^{-itna} \zeta^n(t) \, dt$$

where $\zeta(t)$ is the characteristic function of $Y$ and $na$ is an integer. Finally, if we write $\psi(t) = e^{-iat}\zeta(t)$ for the characteristic function of $Y - a$, we have $\Pr(Y_1 + \cdots + Y_n = na) = (1/2\pi) \int_{-\pi}^{\pi} \psi^n(t) \, dt$.

To evaluate this integral, let us first take it over the range $|t| \leqq \log n/\sqrt{n}$. If we make the usual expansion of $\log \psi(t)$ in terms of the cumulants $\kappa_r$ of $Y_1 - a$, observe $\kappa_1 = 0$, and write $\kappa_2 = \sigma^2$, we find

$$\psi^n(t) = e^{-\frac{n\sigma^2 t^2}{2}} \exp\left\{ n \sum_{r=3}^{6} \frac{\kappa_r (it)^r}{r!} + o(n^{-2}) \right\}$$

when $|t| \leqq \log n/\sqrt{n}$. The transformation $\sqrt{n}\,\sigma\, t = u$ and series expansion of the second factor puts the integrand in the form

$$e^{-u^2/2}\left\{ 1 - \frac{i\kappa_3 u^3}{6\sigma^3\sqrt{n}} + \frac{1}{n}\left[\frac{\kappa_4 u^4}{24\sigma^4} - \frac{\kappa_3^2 u^6}{72\sigma^6}\right] + \frac{uP_1}{n^{3/2}} + \frac{P_2}{n_2} + o(n^{-2})\right\} \tag{2}$$

over $|u| \leqq \sigma \log n$, where $P_i$ denotes a polynomial in $u^2$. Using the fact that

$$\int_{-\sigma \log n}^{\sigma \log n} u^p e^{-u^2/2} \, du = 2^{\frac{p+1}{2}}\, \Gamma\left(\frac{p+1}{2}\right) + o(n^{-2}) \tag{3}$$

when $p$ is even, and vanishes when $p$ is odd, we find

$$\frac{1}{2\pi} \int_{-\log n/\sqrt{n}}^{\log n/\sqrt{n}} \psi^n(t) \, dt = \frac{1}{\sigma\sqrt{2\pi n}} \left\{1 + \frac{1}{8n}\left[\frac{\mu_4}{\mu_2^2} - 3 - \frac{5\mu_3^2}{3\mu_2^3}\right] + o(n^{-2})\right\} \tag{4}$$

where we have expressed the cumulants in terms of the central moments $\mu_r$.

Turning now to the range $\log n/\sqrt{n} \leqq |t| \leqq \pi$, we shall show that this part of the integral is negligible. Since $\kappa_1 = 0$ and $0 < \sigma^2 < \infty$, we can find $0 < t_0 < \pi$ such that $|\psi(t)| \leqq 1 - (\sigma^2 t^2/3)$ for $|t| \leqq t_0$. Therefore, over the range $\log n/\sqrt{n} \leqq |t| \leqq t_0$,

$$\left| \int \psi^n(t) dt \right| \leqq 2 \int_{\log n/\sqrt{n}}^{\infty} e^{-n\sigma^2 t^2/3} \, dt,$$

which is $o(n^{-k})$ for all $k$. As for $t_0 \leqq |t| \leqq \pi$, note first that our assumption that the possible values of $X_1$ are integers whose differences have g.c.d. 1 implies that, when $0 < |t| \leqq \pi$, the points $e^{itx}$ can never all coincide, and hence that $\sum_x q(x)e^{itx}$ lies inside the unit circle. Therefore

$$|\psi(t)| = |e^{-ita} \sum_x q(x)e^{itx}| < 1 \quad \text{for} \quad t_0 \leqq |t| \leqq \pi,$$

and by the continuity of $\psi$ there is a number $\rho < 1$ for which $|\psi(t)| < \rho$ in this range, over which $\int \psi^n(t)\,dt$ is $o(\rho^n)$. We may therefore take the right side of (4) as an expression for $(1/2\pi)\int_{-\pi}^{\pi} \psi^n(t)\,dt$, and hence for

$$\Pr(Y_1 + \cdots + Y_n = na).$$

This fact, combined with Theorem 1, proves Theorem 3.

We present in Table 1 a few illustrations of the accuracy of the two approximations. Here, by the relative error of an approximation $\pi'$ for a quantity $\pi$ we mean $(\pi'/\pi) - 1$. The values of $X$ are $0, 1, \cdots, r - 1$.

**5. The tail probability.** An extension of the methods used above provides expressions for the tail probability. We are indebted to D. A. Darling for suggestions which led to this result.

THEOREM 4: *If $X_1, X_2, \cdots$ are integer-valued random variables satisfying the hypotheses of Theorem 1, then the approximation*

$$\Pi_n^*(a) = \pi_n^*(a)/(1 - z)$$

*for $\Pi_n(a) = \Pr\{X_1 + \cdots + X_n \geqq na\}$ has relative error of order $n^{-1}$, while the approximation*

$$\Pi_n^{**}(a) = \Pi_n^*(a)\left\{1 - \frac{1}{2n}\left[\frac{(z\mu_3/\mu_2) + z(1 + z)/(1 - z)}{(1 - z)\mu_2}\right]\right\}$$

*for $\Pi_n(a)$ has relative error of order $n^{-2}$.*

PROOF: An easy modification of Theorem 1 shows that, for any integer $k$,

$$\begin{aligned}\pi(k) &= \Pr(X_1 + \cdots + X_n = na + k)\\ &= [m(a)]^n e^{-kt^*} \Pr(Y_1 + \cdots + Y_n = na + k)\end{aligned}$$

TABLE 1

| $r$ | $p$ | $a$ | $n$ | $\pi_n$ | Relative error of $\pi_n^*$ | Relative error of $\pi_n^{**}$ |
|---|---|---|---|---|---|---|
| 3 | $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ | $\frac{3}{2}$ | 8 | .040542600 | .0327 | $-.0^3173$ |
| | | | 16 | $.0^260067293$ | .0162 | $.0^4252$ |
| | | | 32 | $.0^318201692$ | $0.0^2804$ | $.0^5740$ |
| | | | 64 | $.0^623350658$ | $0.0^2400$ | $.0^5203$ |
| | | $\frac{7}{4}$ | 8 | $.0^254869684$ | .0556 | $-.0^3816$ |
| | | | 16 | $.0^484094675$ | .0275 | $.0^4654$ |
| | | | 32 | $.0^726832893$ | .0101 | $.0^4192$ |
| | $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ | $\frac{3}{2}$ | 8 | $.0^279040527$ | .0331 | $-.0^2379$ |
| | | | 16 | $.0^324029752$ | .0182 | $-.0^4188$ |
| | | | 32 | $.0^630784454$ | $.0^2901$ | $.0^5307$ |
| 4 | $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ | $\frac{9}{4}$ | 8 | .021789551 | .0291 | $.0^3395$ |
| | | | 16 | $.0^223290971$ | .0142 | $.0^3109$ |

while the proof of Theorem 3 gives

$$\Pr(Y_1 + \cdots + Y_n = na + k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \psi^n(t)\, dt.$$

Summation over $k$ now gives

$$\Pi_n = \lim_{K\to\infty} \sum_{k=0}^{K} \pi(k) = \frac{[m(a)]^n}{2\pi} \lim_{K\to\infty} \int_{-\pi}^{\pi} \frac{1 - e^{-K(it+t^*)}}{1 - e^{-(it+t^*)}} \psi^n(t)\, dt.$$

Because of the boundedness of the integrand, we may pass to the limit inside the integral to get

$$\Pi_n = \frac{[m(a)]^n}{2\pi} \int_{-\pi}^{\pi} \frac{1}{1 - ze^{-it}} \psi^n(t)\, dt. \tag{5}$$

where $z = e^{-t^*} < 1$.

The evaluation of this integral is much like that in the proof of Theorem 3. Since $1/(1 - ze^{-it})$ is bounded, the integral over $|t| \geqq \log n/\sqrt{n}$ is negligible as before. As before, we substitute $\sqrt{n}\,\sigma\, t = u$, and find that when $|u| \leqq \sigma \log n$,

$$\frac{1}{1 - ze^{-it}} = \frac{1}{1-z} - \frac{izu}{\sigma(1-z)^2\sqrt{n}} - \frac{z(1+z)u^2}{2\sigma^2(1-z)^3 n} + \frac{uP_3}{n^{3/2}} + \frac{P_4}{n^2} + o(n^{-2})$$

where the $P_i$ again denote polynomials in $u^2$. Combining this with (2), and integrating the various terms with the aid of (3), we find

$$\int_{-\pi}^{\pi} \frac{1}{1 - ze^{-it}} \psi^n(t)\, dt = \frac{\sqrt{2\pi}}{1-z} \left\{ 1 + \frac{1}{8n}\left[\frac{\mu_4}{\mu_2^2} - 3 - \frac{5\mu_3^2}{3\mu_2^3}\right] - \frac{z}{2n} \frac{\mu_3(1-z) + \sigma^2(1+z)}{\sigma^4(1-z)^2} + o(n^{-2}) \right\}.$$

This, combined with (5), yields Theorem 4.

We present in Table 2 a few illustrations of the accuracy of the approximations. As in Table 1, the values of $X$ are $0, 1, \cdots, r-1$.

TABLE 2

| $r$ | $p$ | $a$ | $n$ | $\Pi_n$ | Relative error of $\Pi_n^*$ | Relative error of $\Pi_n^{**}$ |
|---|---|---|---|---|---|---|
| 3 | $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ | $\frac{4}{3}$ | 8 | .064471879 | 0.148 | $-.0888$ |
| | | | 16 | $.0^2 99484233$ | 0.0846 | $-.0289$ |
| | | | 32 | $.0^3 30990382$ | 0.0465 | $-.00861$ |
| 3 | $\frac{1}{4}, \frac{3}{8}, \frac{3}{8}$ | $\frac{3}{2}$ | 8 | .011276245 | 0.0862 | $-.0259$ |
| | | | 16 | $.0^3 35039405$ | 0.0474 | $-.0^2 733$ |
| 4 | $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ | $\frac{9}{4}$ | 8 | .040328979 | 0.134 | $-.0705$ |
| | | | 16 | $.0^2 44785112$ | 0.0757 | $-.0224$ |

**6. The nonlattice case.** For nonlattice variables, let us heuristically treat $\pi_n^*(a) = m^n(a)/\sigma(a)\sqrt{2\pi n}$ as an approximation to the (in general nonexistent) density of $X_1 + \cdots + X_n$ at the point $na$, and proceed formally.

$$\begin{aligned}
\Pr\{X_1 + \cdots + X_n \geq na\} &\sim \int_0^\infty \pi_n^*\left(a + \frac{x}{n}\right) dx \\
&= \pi_n^*(a) \int_0^\infty \left[\pi_n^*\left(a + \frac{x}{n}\right) \Big/ \pi_n^*(a)\right] dx \\
&\sim \pi_n^*(a) \int_0^\infty \left[\frac{m\left(a + \dfrac{x}{n}\right)}{m(a)}\right]^n dx \\
&\sim \pi_n^*(a) \int_0^\infty \exp\,[xm'(a)/m(a)]\, dx \\
&= \pi_n^*(a) \int_0^\infty \exp\,[-xt^*(a)]\, dx = \pi_n^*(a)/t^*(a).
\end{aligned}$$

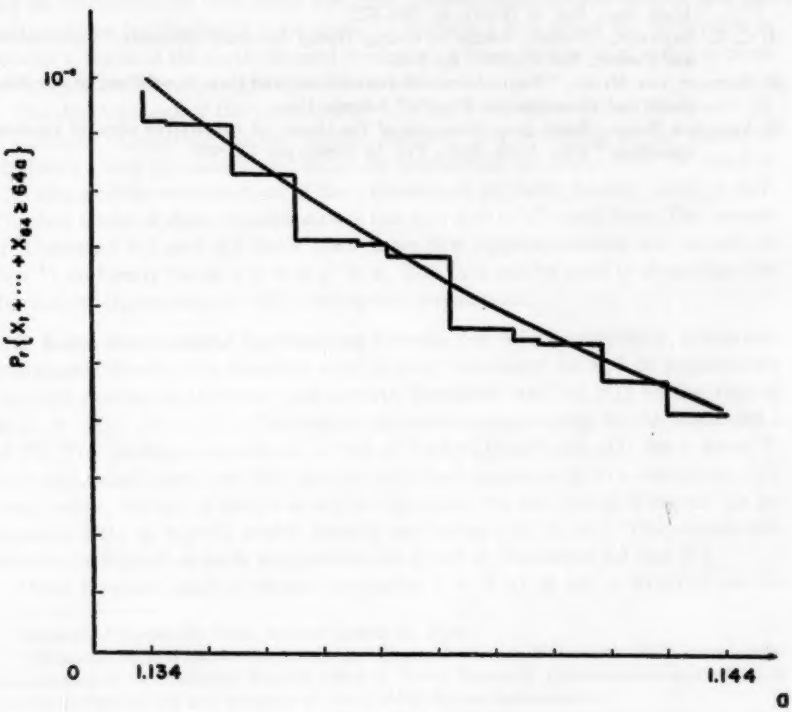


FIG. 1

We thus obtain the approximation

$$\Pi_n^*(a) = \pi_n^*(a)/t^*(a).$$

We conjecture that this approximation has a relative error which is $O(n^{-1})$, just as the corresponding approximation did in the lattice case. For variables with an absolutely continuous component, $\Pi_n^*(a)$ is just the leading term in the expansion obtained by Cramér [2] and is thus known to be correct. The conjecture is supported by numerical evidence for the case in which $X$ has values 0, 1, and $\sqrt{2}$ with equal probabilities. We have computed a portion of the tail of this distribution for $n = 64$, which is shown in Fig. 1 with the approximation superimposed.

## REFERENCES

[1] Herman Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[2] H. Cramér, "Sur un noveau théorème-limite de la théorie des probabilités," *Actualités Scientifiques et Industrielles*, No. 736, Paris, 1938.

[3] W. Feller, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.

[4] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1957), pp. 6–25.

[5] Richard von Mises, "Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik," Leipzig 1931.

[6] Abraham Wald, "Some generalizations of the theory of cumulative sums of random variables," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 287–293.

# BOUNDS ON NORMAL APPROXIMATIONS TO STUDENT'S AND THE CHI-SQUARE DISTRIBUTIONS[1]

By David L. Wallace

*University of Chicago*

**1. Summary.** Formulas closely related to

$$u(t) = [n \log (1 + t^2/n)]^{\frac{1}{2}}$$
$$w(\chi^2) = [\chi^2 - n - n \log (\chi^2/n)]^{\frac{1}{2}}$$

are considered for converting upper tail values of Student's $t$ or chi-square variates with $n$ degrees of freedom to normal deviates. The chief object of the paper is to construct bounds on the deviation from the exact normal deviates such that the absolute deviation is bounded by $cn^{-\frac{1}{2}}$ uniformly in the entire tail. Two approximations for Student's $t$ are suggested that are remarkably accurate and an improvement over other available approximations. The bounds and approximations for Student's $t$ are given in Section 3 and those for chi-square in Section 4. Some of the methods used in obtaining bounds may be of value in other investigations. These are given in Section 2.

The development of the bounds was stimulated by the work of Teichroew [3]. He obtains expansions for the normal deviates corresponding to tail values of Student's $t$ and chi-square and achieves spectacular accuracy even for small $n$. The idea and the construction of the expansion is set forth, briefly, in [4], p. 647. The first terms of these expansions are the $u(t)$ and $w(\chi^2)$ used here. The bounds of Theorems 3.1 and 4.2 show that these first approximations are correct to $O(n^{-\frac{1}{2}})$ uniformly for all $t > 0$ or $\chi^2 > n$. This fact can be used to show that the Teichroew expansions are valid asymptotic expansions.

**2. Some results useful for obtaining bounds.** Let $F$ be an arbitrary, absolutely continuous distribution function with density function $f$, let $\Phi$, $\phi$ be respectively the unit normal distribution and density functions, and let $x(t)$ be the root of $\Phi(x) = F(t)$ (i.e. $x(t)$ is the normal deviate corresponding to the argument $t$ of $F$). The problem considered is that of finding bounds on $x(t)$ for a given $F$. Any numerical bound on $F(t)$ can be converted numerically to a bound on $x(t)$. Frequently, though, a simple analytic expression for the bound is useful. An inequality $F(t) \leqq \Phi(z(t))$ yields directly the bound $x(t) \leqq z(t)$. Two simple sufficient conditions for such inequalities are given in Theorems 2.1 and 2.2.

Often however, only a weaker inequality $1 - F(t) \leqq c[1 - \Phi(z(t))]$ can be

Received October 22, 1958; revised March 31, 1959.

[1] Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I wish to thank Richard Blough and James Yarnold for assistance with the computations.

obtained in which $c$ is typically slightly greater than one (and may depend on $t$). A simple bound on $x(t)$ can still be obtained analytically, although it is not as strong as the one that could be obtained numerically. The bounds are obtained by using the normal tail inequalities and become relatively stronger as $z(t)$ increases. For two directions of inequality, results are given in Theorems 2.3 and 2.4.

*Assume throughout that the density $f(t)$ is positive and continuous for $a < t < \infty$ and that any approximation $z(t)$ to $x(t)$ is a continuously differentiable, strictly increasing function for $a < t < \infty$.* ($a$ is any appropriately chosen constant, which can be $-\infty$ but need not be the lower boundary of the domain of $f$.)

Denote by $g$ the function

$$g(t) = \frac{\phi(z(t))z'(t)}{f(t)}. \tag{2.1}$$

THEOREM 2.1. *If*
(a) $\lim_{t\to\infty} z(t) = \infty$
(b) $\lim_{t\to a} F(t) = \Phi(\lim_{t\to a} z(t))$
(c) $\operatorname{sgn}[g(t) - 1]$ *is a monotonic function of $t$ for $a < t < \infty$, then $x(t) \geqq z(t)$ or $x(t) \leqq z(t)$ for all $a < t < \infty$ according as the function in* (c) *is increasing or decreasing.*

THEOREM 2.2. *If $g(t) \geqq 1/c$ ($\leqq$) for all $a < t < \infty$, then*

$$1 - F(t) \leqq c[1 - \Phi(z(t))]\ (\geqq)\ \textit{for all } a < t < \infty.$$

*If $c = 1$, then $x(t) \geqq z(t)$ ($\leqq$).*

PROOF. Let $\delta(t) = F(t) - \Phi(z(t))$

$$\delta(t) = \int_{z(t)}^{\infty} \phi(u)\,du - \int_t^{\infty} f(s)\,ds.$$

In the first integral, make the substitution $u = z(s)$, so that

$$\delta(t) = \int_t^{\infty} f(s)[g(s) - 1]\,ds.$$

By ($a$) and ($b$), $\delta(a) = 0 = \delta(\infty)$ and if, by (c), $\operatorname{sgn}[g(s) - 1]$ is, say, increasing in $s$, then $\delta(t) \leqq 0$ and $\Phi(z(t)) \geqq F(t) = \Phi(x(t))$ and $z(t) \geqq x(t)$ for all $a < t < \infty$. Theorem 2.2 follows directly from $\delta(t) \geqq [1 - F(t)](1 - c)/c$ ($\leqq$).

Both theorems clearly hold if $\Phi$ is any distribution function with continuous positive density on the entire real line.

THEOREM 2.3. *If, for some value of $t$ such that $z_1(t) > 0$, $F(t)$ satisfies an inequality*

$$1 - F(t) \leqq c_1[1 - \Phi(z_1(t))] \tag{2.2}$$

*with $c_1 \geqq 1$, and if, in addition either* (a) $x(t) > -z_1(t)$ *or* (b) $[1 - \Phi(z_1(t))] \leqq 1/(1 + c_1)$ *holds, then*

$$x(t) \geqq z_1(t) - \frac{c_1 - 1}{z_1(t)}.$$

THEOREM 2.4. *If, for some value of t, such that $z_2(t) > 0$, $F(t)$ satisfies an inequality*

$$1 - F(t) \geqq c_2[1 - \Phi(z_2(t))] \tag{2.3}$$

*with $0 < c_2 \leqq 1$, then*

$$x(t) \leqq z_2(t) + \frac{1 - c_2}{c_2} \frac{1}{z_2(t)}.$$

If $c_1 \leqq 1$ in (2.2) it may be replaced by 1 and the bound $x(t) \geqq z_1(t)$ used. $c_2 \geqq 1$ in (2.3) can be handled similarly. Results taking advantage of these constants can be obtained but are rather poor.

PROOFS. By definition of $x(t)$, $1 - F(t) = 1 - \Phi(x(t))$. Henceforth the argument $t$ in $x(t)$ and $z(t)$ will be dropped. The proofs use the Taylor expansion

$$\Phi(z) = \Phi(x) + (z - x)\phi(\theta z + (1 - \theta)x), \qquad 0 \leqq \theta \leqq 1,$$

and the normal tail inequality

$$1 - \Phi(u) < \frac{\phi(u)}{u}, \qquad u > 0. \tag{2.4}$$

Inequality (2.2) and condition (b) of Theorem 2.3 together imply condition (a) since $1 - \Phi(x) \leqq c_1[1 - \Phi(z_1)] \leqq c_1/(1 + c_1)$ so that $\Phi(x) \geqq 1/(1 + c_1) \geqq 1 - \Phi(z_1) = \Phi(-z_1)$ and hence $x \geqq -z_1$.

Eliminating $\Phi(x)$ between inequality (2.2) and the Taylor expansion and solving for $x$ gives

$$x \geqq z_1 - \frac{(c_1 - 1)[1 - \Phi(z_1)]}{\phi(\theta z_1 + (1 - \theta)x)}. \tag{2.5}$$

Let $c_1 \geqq 1$ and assume first that $x \leqq z_1$ so that, with condition (a), $|x| \leqq z_1$ and hence $\phi(\theta z_1 + (1 - \theta)x) \geqq \phi(z_1)$. Using this and inequality (2.4) in inequality (2.5), $x \geqq z_1 - (c_1 - 1)/z_1$. But this holds trivially when $x \geqq z_1$, so that Theorem 2.3 is proved.

Eliminating $\Phi(z)$ between inequality (2.3) and the Taylor expansion and solving for $x$ gives

$$x \leqq z_2 + \frac{(1 - c_2)}{c_2} \frac{(1 - \Phi(x))}{\phi(\theta z_2 + (1 - \theta)x)}. \tag{2.6}$$

Let $0 < c_2 \leqq 1$, and assume first that $x \geqq z_2$. Then $\phi(\theta z_2 + (1 - \theta)x) \geqq \phi(x)$ and with inequality (2.4),

$$x \leqq z_2 + \frac{(1 - c_2)}{c_2 |x|} \leqq z_2 + \frac{(1 - c_2)}{c_2 z_2}.$$

These inequalities hold trivially when $x \leqq z_2$ and Theorem 2.4 is proved.

Let $\{F_n(t)\}$ be a sequence of distribution functions and $\{x_n(t)\}$ the corresponding normal deviates. An approximate normal deviate $z_n(t)$ which is a close approximation to $x_n(t)$ in the entire tail of the distribution would often be useful.

The results of this section enable detailed boundings of the errors of such approximations from the corresponding distribution function approximations. The essential qualitative result is that the absolute deviate error will be of order $\rho(n)$ throughout the entire tail if the per cent error (relative to smaller tail) in the distribution function approximation is of order $\rho(n)$ throughout the tail. The result is not quite necessary.

**3. Normal approximations to Student's $t$ distribution.** Let $F_n$ be the distribution function of Student's $t$ on $n$ degrees of freedom.

$$1 - F_n(t) = a_n(2\pi)^{-\frac{1}{2}} \int_t^\infty \left(1 + \frac{s^2}{n}\right)^{-\frac{n+1}{2}} ds,$$

$$a_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{2}{n}\right)^{\frac{1}{2}}.$$

Denote by $x_n(t)$ the normal deviate corresponding to the deviate $t$ of Student's distribution. Chu [1] has studied the normal approximation $\Phi(t)$ of $F(t)$. He was not concerned with approximations in the extreme tails of the distribution nor with quantile approximations; but methods similar to his can be used.

Bounds on the deviate $x_n(t)$ are given by

THEOREM 3.1. *For all $t > 0$ and with $u(t) = [n \log (1 + t^2/n)]^{\frac{1}{2}}$ and $k = .368$,*

(a) $x_n(t) \leqq u(t)$, $\quad n > 0$;

(b) $x_n(t) \geqq u(t)(1 - (1/2n))^{\frac{1}{2}} \equiv u_2(t)$, $\quad n > .50$;

(c) $x_n(t) \geqq u(t) - k/n^{\frac{1}{2}} \equiv u_3(t)$, $\quad n \geqq .50$.

COROLLARY. *Inequality* (b) *can be written as*

(b') $x_n(t) \geqq u(t)(1 - b_1/n))$, $\quad n \geqq n_0 > .50$,

*with $b_1 = n_0[1 - (1 - 1/2n_0)^{\frac{1}{2}}]$. Three numerical values of $b_1$ which will suffice for almost all uses are: $n_0 = 1, b_1 = .293$; $n_0 = 3, b_1 = .262$; $n_0 = 10, b_1 = .254$.*

The bounds show that $u(t)$, as an approximation to $x_n(t)$, has an absolute error not exceeding $.368n^{-\frac{1}{2}}$ and a relative error (relative to $u(t)$) not exceeding $b_1/n$. Except for very large values of $t$, the bound (c) is much poorer than the bound (b). The main interest in (c) is the rather remarkable fact that even as $t$ and $x_n(t)$ increase indefinitely the error remains bounded and even of order $n^{-\frac{1}{2}}$. An interesting theoretical application will be noted in Section 6.

The derivations of the bounds and a few calculations suggest the following conjectures on the behavior of $x_n(t)$: that $x_n(t)/u_2(t) \to 1$ as $t \to 0$, and that $u(t) - x_n(t)$ as a function of $t$, increases monotonically to a maximum value slightly less than $.368n^{-\frac{1}{2}}$ and then decreases monotonically to zero, the maximum occurring for $t$ and $n$ for which $u^2(t)/n$ is substantial.

Calculations indicate that the error, $u(t) - x_n(t)$, is close to its maximum value

unless $t$ is very large, so that the maximum of the two bounds (b) and (c) is a good approximation. Two superior approximations that were obtained empirically are *approximations* $u_4(t)$ *and* $u_5(t)$,

$$u_4(t) = u(t)\left(\frac{8n+1}{8n+3}\right),$$

$$u_5(t) = u(t) - \frac{2u(t)}{8n+3}\,(1 - e^{-s^2})^{\frac{1}{2}}$$

*with*

$$s = \frac{.368(8n+3)}{2\sqrt{n}\,u(t)}.$$

*For all* $n > 1$ *and all* $t > 0$,

$$u_2(t) < u_4(t) < u(t),$$

$$\max\,(u_2(t), u_3(t)) < u_5(t) < u(t).$$

$u_4$ was chosen as slightly larger than $u_2$ to give a good fit for small $t^2/n$. $u_5$ was constructed to be larger than $u_4$ and $u_3$ and to so join them as to give excellent approximation over a wide range of values. Though the function is somewhat complicated, it is amenable to slide rule calculation. $u_4$ seems to be within .02 of $x$ for $t^2/n$ less than about 5 and $u_5$ within .02 of $x$ for a much wider range.

The bounds $u(t)$, $u_2(t)$, $u_3(t)$, the approximations $u_4(t)$, $u_5(t)$, and the approximation $u_6(t)$ obtained from the Paulson approximation [2] to $F$ are illustrated in Table 1 for $n = 1, 3, 10$, and selected values of $t$. The Paulson approximation gives a normal deviate corresponding to the double tail $t$ probability and hence has to be converted to be comparable.

$$K_p(t) = \frac{\left[\frac{2}{9n}\,t^{2/3} - \frac{2}{9}\right]}{\left[\frac{2}{9n}\,t^{4/3} + \frac{2}{9}\right]^{\frac{1}{2}}}, \qquad 1 - \Phi(u_6(t)) = \frac{1}{2}[1 - \Phi(K_p(t))]$$

Polynomial approximations such as the Hotelling-Frankel approximations, are very poor for small $n$ or for very large $t$.

All bounds and approximations except $u_5(t)$ can easily be inverted analytically to give bounds or approximations for the Student's deviate corresponding to a given normal deviate, i.e., for the quantiles of $t$.

The proof of the theorem will be preceded by two lemmas.

Lemma 1. *For all* $x > 0$, $h_c(x) = (e^x - 1)/xe^{cx}$ *is monotone decreasing for* $c = 1$, *monotone increasing for* $c = \frac{1}{2}$ *and not monotonic for* $\frac{1}{2} < c < 1$.

Proof. $h_c'(x) = (1/x^2e^{cx})[xe^x - (e^x - 1)(cx + 1)]$ and is $\geqq 0$ or $\leqq 0$ according as $xe^x/((e^x - 1)(cx + 1))$ is $\geqq 1$ or $\leqq 1$. The result follows from a termwise comparison of the Maclaurin expansions of the numerator and denominator.

TABLE 1

*Bounds on the normal deviate $x_n(t)$ for Student's distribution*

$$1 - \Phi(x_n(t)) = 1 - F_n(t)$$

| $n$ | $t$ | Exact $x_n(t)$ | Bounds from Theorem 3.1 | | | Approximation | | |
|---|---|---|---|---|---|---|---|---|
| | | | Upper $u(t)$ | Lower $u_2(t)$ | Lower $u_3(t)$ | $u_4(t)$ | $u_5(t)$ | $u_6(t)$ |
| 1 | 0.3 | .235 | .294 | .208 | <0 | .241 | .241 | .257 |
| | 1 | .674 | .832 | .589 | .465 | .680 | .681 | .674 |
| | 2 | 1.047 | 1.269 | .897 | .901 | 1.038 | 1.048 | 1.031 |
| | 4 | 1.419 | 1.683 | 1.190 | 1.315 | 1.377 | 1.416 | 1.349 |
| | 8 | 1.756 | 2.043 | 1.445 | 1.675 | 1.672 | 1.750 | 1.576 |
| | 12 | 1.935 | 2.231 | 1.577 | 1.863 | 1.825 | 1.927 | 1.670 |
| | $10^2$ | 2.729 | 3.035 | 2.146 | 2.667 | 2.177 | 2.704 | 1.896 |
| | $10^6$ | 4.514 | 4.799 | 3.393 | 4.431 | 3.926 | 4.447 | 1.964 |
| 3 | 1 | .858 | .929 | .848 | .717 | .860 | .860 | .855 |
| | 2 | 1.478 | 1.594 | 1.455 | 1.382 | 1.476 | 1.478 | 1.477 |
| | 4 | 2.197 | 2.353 | 2.148 | 2.141 | 2.179 | 2.197 | 2.160 |
| | 8 | 2.872 | 3.053 | 2.787 | 2.840 | 2.826 | 2.879 | 2.705 |
| | 12 | 3.228 | 3.417 | 3.119 | 3.204 | 3.164 | 3.237 | 2.935 |
| | $\sqrt{3} \times 10^2$ | 5.057 | 5.256 | 4.797 | 5.044 | 4.866 | 5.058 | 3.493 |
| 10 | 1 | .952 | .976 | .952 | .860 | .953 | .953 | .948 |
| | 2 | 1.790 | 1.834 | 1.788 | 1.718 | 1.790 | 1.790 | 1.805 |
| | 4 | 3.021 | 3.091 | 3.013 | 2.975 | 3.017 | 3.020 | 3.014 |
| | 8 | 4.382 | 4.474 | 4.361 | 4.357 | 4.366 | 4.384 | 4.279 |
| | 12 | 5.128 | 5.229 | 5.097 | 5.113 | 5.103 | 5.133 | 4.902 |
| 100 | 100 | 21.447 | 21.483 | 21.429 | 21.446 | 21.429 | 21.450 | 18.541 |

LEMMA 2. *For all $x > 0$, $((e^x - 1)e^{2kx^{1/2}})/xe^x \geqq 1$ with $k = .368$.*

PROOF. The desired inequality is equivalent to the inequality

$$Q(x) = e^x - 1 - xe^{x-2kx^{1/2}} \geqq 0.$$

Let $T$ be defined by

$$Q'(x) = e^{x-2kx^{1/2}}[e^{2kx^{1/2}} - (1 - kx^{1/2} + x)] = e^{x-2kx^{1/2}}T(x).$$

The simultaneous equations in $x$ and $k$: $Q(x) = 0$ and $T(x) = 0$ will have exactly one solution with positive $x$ and the root for $k$ is (to three decimals) the smallest value for which the inequality $Q(x) \geqq 0$ holds for all $x > 0$. The solution is $k = .368$ and $x = 7.312$.

PROOF OF THEOREM. Proceeding as in Theorem 2.1, set $z(t) = \lambda u(t) - \mu$ with $u(t) = [n \log (1 + t^2/n)]^{1/2}$ and with $\lambda$, $\mu$ constants to be chosen. Then form the function $g(t) = \phi(z(t))z'(t)/f_n(t)$ written as a function of $x = u^2/n$, which is monotonic in $t$,

$$g^2(t) = h(x) = \frac{\lambda^2 e^{-\mu^2}}{a_n^2} \frac{(e^x - 1)e^{2\lambda\mu(nx)^{1/2}}}{xe^{cx}}$$

where $c = 1 - n(1 - \lambda^2)$.

First, set $\mu = 0$. Then $z(t) = \lambda u(t)$ satisfies conditions (a), (b) of Theorem 2.1. Monotony of $g(t)$ and hence condition (c) of Theorem 2.1 follow from Lemma 1: decreasing for $c = 1$ or $\lambda = 1$, increasing for $c = \frac{1}{2}$ or $\lambda = (1 - (1/2n))^{\frac{1}{2}}$. Conclusions (a) and (b) follow.

Next set $\lambda = 1$ and $\mu = k/n^{\frac{1}{2}}$ with $k = .368$. Then, using Lemma 2, $g(t) \geqq 1$ for all $t > 0$, provided that $a_n e^{k^2/2n} \leqq 1$. Hence (c) follows from Theorem 2.

The proof that $a_n e^{k^2/2n} \leqq 1$ for all $n \geqq .50$ and that $(1 - (1/2n))^{\frac{1}{2}} \geqq 1 - b/n$ from which the Corollary follows are given in Section 5.

**4. Normal approximations to the chi-square distribution.** Let $F_n$ be the distribution function of chi-square on $n$ degrees of freedom (using $t$ instead of $\chi^2$ as argument).

$$1 - F_n(t) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_t^\infty s^{\frac{n}{2}-1} e^{-s/2}\, ds.$$

Denote by $y_n(t)$ the normal deviate corresponding to the chi-square argument $t$. Only the upper tail with $t > n$ is treated in this paper. Bounds on $1 - F_n(t)$ and $y_n(t)$ are given by

THEOREM 4.1. *For all $t > n$, all $n > 0$, and with $w(t) = [t - n - n \log (t/n)]^{\frac{1}{2}}$, and $w_2(t) = w(t) + \frac{1}{3}(2/n)^{\frac{1}{2}}$*

(a) $1 - F_n(t) > d_n e^{1/9n}[1 - \Phi(w_2(t))]$

(b) $1 - F_n(t) < d_n[1 - \Phi(w(t))]$

*in which*

$$d_n = \frac{\left(\frac{n}{2}\right)^{\frac{n-1}{2}} e^{-\frac{n}{2}} (2\pi)^{\frac{1}{2}}}{\Gamma(n/2)}.$$

THEOREM 4.2. *For all $t > n$,*

(a) $y_n(t) \leqq w_2(t) + (1/w_2(t)) \max [0, d_n^{-1}e^{-1/9n} - 1]$, $\quad n > 0$,

(b) $y_n(t) \geqq w(t)$, $\quad n > .37$.

COROLLARY 1. *Inequality* (a) *can be written as*

(a′) $y_n(t) \leqq w_2(t) + b_2/nw_2(t)$, $\quad n \geqq n_0 > 0$,

*with* $b_2 = n_0(e^{1/18n_0} - 1)$. Numerical values of $b_2$ which will suffice for almost all uses are: $n_0 = .37$, $b_2 = .060$; $n_0 = 1$, $b_2 = .058$, $n_0 = 10$, $b_2 = .056$.

COROLLARY 2. *For all $t > n$ and all $n > .37$,*

$$w(t) \leqq y_n(t) \leqq w(t) + .60n^{\frac{1}{2}}.$$

The bounds on $y_n(t)$ are illustrated in Table 2 for $n = 8$ and selected values of

TABLE 2

*Bounds on the normal deviate $y_n(t)$ for the chi-square distribution*

$$1 - \Phi(y_n(t)) = 1 - F_n(t), \qquad n = 8$$

| $t$ | $\frac{t-n}{(2n)^{1/2}}$ | Exact $y_n(t)$ | Bounds from Theorem 4.2 | | Wilson-Hilferty $w_3(t)$ |
|---|---|---|---|---|---|
| | | | Upper ($a'$) | Lower ($b$) | |
| 12 | 1 | 1.031 | 1.095 | .869 | 1.035 |
| 16 | 2 | 1.724 | 1.769 | 1.566 | 1.726 |
| 20 | 3 | 2.314 | 2.354 | 2.160 | 2.310 |
| 24 | 4 | 2.835 | 2.874 | 2.685 | 2.820 |
| 32 | 6 | 3.737 | 3.776 | 3.593 | 3.691 |
| 40 | 8 | 4.512 | 4.553 | 4.373 | 4.427 |
| 72 | 16 | 6.940 | 6.989 | 6.813 | 6.647 |

$t$. Shown are bounds (a) and (b), the exact normal deviate $y_n(t)$ and the Wilson-Hilferty [6] approximate deviate

$$w_3(t) = \frac{(t/n)^{1/3} - 1 + \dfrac{2}{9n}}{\left(\dfrac{2}{9n}\right)^{\frac{1}{2}}}.$$

The Wilson-Hilferty approximation is much superior to the bounds as approximations except in the extreme tail and the chief value of the approximation is the uniform bound of order $n^{-\frac{1}{2}}$ on the error in the tail.

The proof of the theorem will be preceded by a lemma.

LEMMA 3. *For all* $x > 0$,

(a) $\lambda(x) < x$

(b) $e^{\frac{1}{3}\lambda(x)}\lambda(x) > x$

*with* $\lambda(x) = 2^{\frac{1}{2}}[x - \log(1 + x)]^{\frac{1}{2}}$.

Inequality (a) follows immediately from $\log(1 + x) > x - x^2/2$. It cannot be improved by any factor of the form $\exp(k\lambda(x))$.

Inequality (b) is sharp for small $x$ and the coefficient in the exponent cannot be decreased. Let

$$y_1 = e^{2\lambda(x)/3}, \qquad y_2 = x^2/u, \qquad u = \lambda^2(x).$$

Denote derivatives with respect to $x$ by primes. The proof consists in showing that $y_1' > \frac{2}{3}$ and $y_2' < \frac{2}{3}$ for all $x > 0$, from which it follows that $y_2 < 1 + 2x/3 < y_1$ and hence, inequality ($b$): $y_1^{\frac{1}{2}} > y_2^{\frac{1}{2}}$.

$$2/3 - y_2' = \frac{2}{3u^2}\left[u^2 - 3ux + \frac{3x^3}{1 + x}\right] = \frac{2}{3u^2}\,\beta(x),$$

$$\beta(x) = 4\left[\log(1 + x) - \frac{x}{4}\right]^2 + \frac{3x^2(x - 3)}{4(1 + x)}.$$

Hence $\beta(x) > 0$ for all $x \geqq 3$.

$$\beta'(x) = \frac{2(3-x)}{1+x}\left[\log(1+x) - \frac{x}{4}\right] + \frac{3x[x^2-3]}{2(1+x)^2}.$$

Let

$$\gamma(x) = \frac{(1+x)}{2(3-x)}\beta'(x) = \log(1+x) - x + \frac{3x^2}{2(1+x)(3-x)}.$$

$$\gamma'(x) = \frac{x^3(5-x)}{(1+x)^2(3-x)^2}.$$

Hence $\gamma'(x) > 0$ for all $0 < x < 5$. $\gamma(0) = 0$ so that $\gamma(x) > 0$ for all $0 < x < 5$. Then $\beta'(x) > 0$ for all $0 < x < 3$. Since $\beta(0) = 0$, $\beta(x) > 0$ for $0 < x < 3$, which, combined with the result for $x \geqq 3$ gives $y_2' < \frac{2}{3}$ and $y_2 < 1 + 2x/3$ for all $x > 0$.

Let

$$\delta(x) = \frac{3\lambda^3(1+x)^2 y_1''}{2y_1}.$$

Then $\delta(x) = \lambda^2 - x^2 + (\frac{2}{3})x^2\lambda = \lambda^2(1 - y_2) + (\frac{2}{3})x^2\lambda$. Using the inequalities $y_2 < 1 + 2x/3$ and $\lambda \leqq x$ gives $\delta(x) > 0$ for all $x > 0$. Since $y_1'(0) = \frac{2}{3}$ and $y_1(0) = 1$, the desired result $y_1 > 1 + 2x/3$ for $x > 0$ follows immediately and inequality (b) is proved.

Proof of Theorem. Set $z(t) = w(t) + c(2/n)^{\frac{1}{2}}$ and form the function $g(t)$ of (2.1), written as a function of $x = (t-n)/n$, then

$$g(t) = d_n^{-1} e^{-c^2/n} \frac{xe^{-c\lambda(x)}}{\lambda(x)}$$

with $\lambda(x) = [2(x - \log(1+x))]^{\frac{1}{2}}$. Using Lemma 3 and Theorem 2.2, with $c$ equal to 0 and $\frac{1}{3}$, Theorem 4.1 follows.

The first part of Theorem 4.2 follows using Theorem 2.4 and the second part from the fact, proved in Section 5, that $d_n < 1$ for $n \geqq .37$. Corollary 1 follows from the fact, proved in Section 5, that $e^{-(1/9n)}d_n^{-1} < 1 + b_2/n$ for $n \geqq n_0 > 0$. Corollary 2 follows from Corollary 1 and the theorem.

**5. Bounding of some simple functions.** In this section four results, used in Sections 3 and 4, are derived. Specifically, with

$$a_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{2}{n}\right)^{\frac{1}{2}}, \qquad d_n = \frac{\left(\frac{n}{2}\right)^{\frac{n-1}{2}} e^{-\frac{n}{2}} (2\pi)^{\frac{1}{2}}}{\Gamma(n/2)}, \qquad k = .368,$$

$$b_1 = n_0\left[1 - \left(1 - \frac{1}{2n_0}\right)^{\frac{1}{2}}\right], \qquad b_2 = n_0\,[e^{1/18n_0} - 1],$$

$$a_n e^{k^2/2n} \leqq 1, \qquad n \geqq .50; \tag{5.1}$$

$$\left(1 - \frac{1}{2n}\right)^{\frac{1}{2}} \geqq 1 - \frac{b_1}{n}, \qquad n \geqq n_0 > .50; \tag{5.2}$$

$$d_n \leqq 1, \qquad n \geqq .37; \tag{5.3}$$

$$e^{-\frac{1}{9n}} d_n^{-1} \leqq 1 + \frac{b_2}{n}, \qquad n \geqq n_0 > 0. \tag{5.4}$$

An easily proved result that is used repeatedly (with $x = 1/n$) is the following:

LEMMA 4. *If $f(x)$ has a uniformly convergent Maclaurin series for $0 \leqq x \leqq x_0$ and if all derivatives of $f(x)$ at $x = 0$ of order greater than $m$ are of constant sign, say positive, then for all $0 \leqq x \leqq x_0$,*

$$T_m(x) \leqq f(x) \leqq T_{m-1}(x) + x^m \left[\frac{f(x_0) - T_{m-1}(x_0)}{x_0^m}\right]$$

*where $T_m(x)$ is the partial sum through order $m$ of the Maclaurin series. (If sign is negative, the direction of the inequalities is reversed.)*

(5.2) is a direct application of Lemma 4.

The Stirling expansion with argument $n/2$ is just the expansion of $-\log d_n$ and the first two partial sums bracket the value ([5], p. 253).

$$\frac{1}{6n} - \frac{1}{45n^3} \leqq -\log d_n \leqq \frac{1}{6n}, \qquad n > 0. \tag{5.5}$$

By the duplication formula for the gamma function, $a_n = d_n^2/d_{2n}$ so that,

$$-\frac{1}{4n} - \frac{1}{360n^3} \leqq \log a_n \leqq -\frac{1}{4n} + \frac{2}{45n^3}, \qquad n > 0. \tag{5.6}$$

Using (5.5), it follows that $\log d_n \leqq 0$ and hence (5.3) for $n^2 \geqq 2/15$ or $n \geqq .37$. Also, for all $n > 0$,

$$d_n^{-1} e^{-(1/9n)} - 1 \leqq e^{1/18n} - 1$$

and (5.4) follows by application of Lemma 4.

From (5.6) it follows that

$$a_n\, e^{k^2/2n} \leqq \exp\left[\frac{1}{n}\left(-\frac{1}{4} + \frac{k^2}{2} + \frac{2}{45n^2}\right)\right].$$

The exponent is negative if $n \geqq .494$ proving (5.1).

## REFERENCES

[1] J. T. CHU, "Errors in normal approximations to the $t$, $\tau$, and similar types of distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 780–789.

[2] EDWARD PAULSON, "An approximate normalization of the analysis of variance distribution," *Ann. Math. Stat.*, Vol. 13 (1942), p. 233.

[3] D. TEICHROEW, "Generalized normalization polynomials for the $t$ and incomplete gamma distributions," personal communication.

[4] DAVID L. WALLACE, "Asymptotic approximations to distributions," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 635–654.

[5] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, American ed., Macmillan and Cambridge University Press, New York, 1943.

[6] E. B. WILSON AND M. M. HILFERTY, "The distribution of chi-square," *Proc. Nat. Acad. of Sci. USA*, Vol. 17 (1931), pp. 684–688.

# APPROXIMATE EXPRESSIONS FOR THE CONDITIONAL MEAN AND VARIANCE OVER SMALL INTERVALS OF A CONTINUOUS DISTRIBUTION

By Gunnar Ekman

*Institute of Mathematical Statistics, University of Stockholm*

**1. Summary.** Approximate expressions of more or less simple analytical form are derived for the conditional mean and variance over small intervals of a distribution having a probability density of somewhat restricted nature. An alternative formula for the mean is derived. This result is applied to an extremal problem in stratified sampling.

**2. Derivation of results.** Consider a positive function $f(t)$, defined on some finite interval containing points $x$ and $y$, and having continuous derivatives to the fourth order. Let us define functions

$$(1)\qquad I_i(y, x) = \int_y^x (x - t)^i f(t)\, dt, \qquad \begin{cases} i = 0, 1, 2 \\ a \leqq x, y \leqq b. \end{cases}$$

These functions exist and may be partially integrated, whereby, using the mean value theorem for integrals and writing $(x - y) = k$ for ease of notation, the following identities are obtained:

$$(2)\qquad \begin{aligned} I_0(y, x) &= kf + \frac{k^2}{2!}f' + \frac{k^3}{3!}f'' + \frac{k^4}{4!}f''' + \frac{k^5}{5!}f^{(4)}, \\ I_1(y, x) &= \frac{k^2}{2!}f + \frac{k^3}{3!}f' + \frac{k^4}{4!}f'' + \frac{k^5}{5!}f''' + \frac{k^6}{6!}f^{(4)}, \\ I_2(y, x) &= 2\left\{\frac{k^3}{3!}f + \frac{k^4}{4!}f' + \frac{k^5}{5!}f'' + \frac{k^6}{6!}f''' + \frac{k^7}{7!}f^{(4)}\right\}, \end{aligned}$$

where all $f^{(\nu)}$ but $f^{(4)}$ are taken at the point $y$, $f^{(4)}$ being taken at points $\theta_i$ within $(x, y)$. Let us now define three new functions by

$$(3)\qquad \begin{aligned} H_1(y, x) = \frac{I_1(y, x)}{I_0(y, x)} = \frac{k}{2}\Bigg\{1 - \frac{1}{6}\bigg[k \cdot \frac{f'}{f} &+ \frac{k^2}{2!}\left(\frac{ff'' - (f')^2}{f^2}\right) \\ &+ \frac{k^3}{3!}\left(\frac{9f'''}{10f} + \frac{3(f')^3}{2f^3} - \frac{5f'f''}{2f^2}\right)\bigg] + O(k^4)\Bigg\}, \end{aligned}$$

$$(4)\qquad \begin{aligned} H_2(y, x) = \frac{I_2(y, x)}{I_0(y, x)} = \frac{k^2}{3}\Bigg\{1 - \frac{k}{4} \cdot \frac{f'}{f} &- k^2\left(\frac{7f''}{60f} - \frac{(f')^2}{8f^2}\right) \\ &- k^3\left(\frac{f'''}{30f} + \frac{(f')^3}{16f^3} - \frac{f'f''}{10f^2}\right) + O(k^4)\Bigg\}, \end{aligned}$$

Received December 1, 1958; revised April 13, 1959.

$$
(5)\qquad H_3(y,x) = H_2(y,x) - [H_1(y,x)]^2 = \frac{k^2}{12}\left\{1 + k^2\left(\frac{f''}{30f} - \frac{(f')^2}{12f^2}\right) + k^3\left(\frac{f'''}{60f} + \frac{(f')^3}{12f^3} - \frac{f'f''}{10f^2}\right) + O(k^4)\right\},
$$

where we have imposed yet another condition on $f(t)$, namely $f(y) \neq 0$. Furthermore we have the Taylor expansion

$$
(6)\qquad \log f(x) - \log f(y) = k \cdot \frac{f'}{f} + \frac{k^2}{2!}\left(\frac{ff'' - (f')^2}{f^2}\right) + \frac{k^3}{3!}\left(\frac{f'''}{f} + \frac{2(f')^3}{f^3} - \frac{3f'f''}{f^2}\right) + O(k^4).
$$

Using (6), we have from (3) and (5)

$$
(7)\qquad \left[H_1(y,x) - \frac{k}{2} + \frac{k}{12}\cdot \log\frac{f(x)}{f(y)}\right] = \frac{-k^4}{720}\cdot R_1(y) + O(k^5),
$$

$$
(8)\qquad \left[H_3(y,x) - \frac{k^2}{12} + \left(\frac{k}{12}\cdot \log\frac{f(x)}{f(y)}\right)^2\right] = \frac{k^4}{360}\cdot\frac{f''(y)}{f(y)} + \frac{k^5}{720}\cdot R_2(y) + O(k^6),
$$

where

$$
(9)\qquad \begin{aligned} R_1(t) &= -\left[\frac{f'''(t)}{f(t)} + \frac{5(f'(t))^3}{(f(t))^3} - \frac{5f'(t)f''(t)}{(f(t))^2}\right],\\ R_2(t) &= \left[\frac{f'''(t)}{f(t)} - \frac{f'(t)f''(t)}{(f(t))^2}\right]. \end{aligned}
$$

From the definition of $H_1(y, x)$ and $H_3(y, x)$, by (1) and (3)–(5), we see that the functions $[H_1(y, x) - (k/2)]$ and $H_3(y, x)$ are symmetrical in $x$ and $y$, so that the same may be said for the left hand members of (7) and (8). This implies that $x$ and $y$ may be interchanged in the right hand members of these identities without changing the order of magnitude of the terms.

Now we see from (1) and (3)–(5), if $\mu(y, x)$ and $\sigma^2(y, x)$ denote the conditional mean and conditional variance respectively over $(y, x)$ of the function $f(t)$ considered as a probability density, that we have

$$
H_1(y,x) = x - \mu(y,x), \qquad H_3(y,x) = \sigma^2(y,x),
$$

so that we are led from (7) and (8) to the following approximations:

$$
(10)\qquad \mu(y,x) \sim \frac{(x+y)}{2} + c(y,x) + \frac{(x-y)^4}{720}\cdot R_1(y),
$$

$$
(11)\qquad \sigma^2(y,x) \sim \frac{(x-y)^2}{12} - [c(y,x)]^2 + \frac{(x-y)^4}{360}\cdot\frac{f''(y)}{f(y)} + \frac{(x-y)^5}{720}\cdot R_2(y),
$$

where

$$(12) \qquad c(y, x) = \frac{(x - y)}{12} \cdot \log \frac{f(x)}{f(y)},$$

and where $R_1(t)$, $R_2(t)$ are given by (9); $x$ and $y$ may be interchanged in (10) and (11). We note that by taking only values of $f(t)$ at the end points $(x, y)$ into consideration, that is, by neglecting all but the first two terms in (10) and (11), we may approximate $\mu$ and $\sigma^2$ correctly to $O(x - y)^4$. We call attention to the fact that the logarithms are to the base $e$, and that the conditions imposed on $f(t)$ are the following: $0 < f(x), f(y) < \infty$, and the first four derivatives of $f$ exist and are continuous; these conditions may be weakened and were imposed for ease of derivation. Finally, it may be remarked that approximations containing $I_0(y, x)$ explicitly may be derived in the case $f(x)$ or $f(y) = 0$.

**3. Further results.** We shall derive another approximation to $\mu(y, x)$, assuming the existence and continuity of $f'$ and $f''$ and with $f(x), f(y) \neq 0$. By cubing both sides of (3) we find

$$(13) \qquad [x - \mu(y, x)]^3 = \frac{k^3}{8}\left[1 - \frac{k}{2} \cdot \frac{f'}{f} + O(k^2)\right].$$

On the other hand, using the Taylor expansion

$$f(x) = f + k \cdot f' + O(k^2)$$

together with (2), we obtain from (13)

$$(14) \qquad \frac{k^2}{8} \cdot \frac{I_0(y, x)}{f(x)} = \frac{k^3}{8}\left[1 - \frac{k}{2} \cdot \frac{f'}{f} + O(k^2)\right] = [x - \mu(y, x)]^3 + O(k^5).$$

Assuming $x > y$ for definitiveness and writing $I_0(y, x) = P(y, x) =$ the area under $f(t)$ in $(y, x)$, we have from (14)

$$(15) \qquad \mu(y, x) = x - \left\{\frac{P(y, x) \cdot (x - y)^2}{8f(x)}\right\}^{\frac{1}{3}} + O(k^3),$$

from which, by permuting $x$ and $y$, we obtain also

$$(16) \qquad \mu(y, x) = y + \left\{\frac{P(y, x) \cdot (x - y)^2}{8f(y)}\right\}^{\frac{1}{3}} + O(k^3).$$

These approximations, (15) and (16), are less accurate than (10), even when the last term of the latter is neglected, but may be used to obtain an approximate solution to the following problem arising in the theory of stratified sampling with proportionate allocation, see [1]. Given a density $f(x)$ over a range $(x_0, x_n)$, $(n - 1)$ variable points $x_i$, $i = 1, \cdots, (n - 1)$, $x_{i-1} < x_i$, and denoting by $P_h$, $\mu_h$ and $\sigma_h^2$ the area, conditional mean and conditional variance in the interval $(x_{h-1}, x_h)$, we are to minimize

$$(17) \qquad \sum_{h=1}^{n} P_h \sigma_h^2.$$

In [1] it is shown that the points minimizing (17) satisfy

$$x_h - \mu_h = \mu_{h+1} - x_h, \qquad h = 1, \cdots, (n-1), \tag{18}$$

and it is seen that the determination of such points may give rise to some computational difficulties. Let us now assume $(x_n - x_0)$ finite, $0 < f(x) < \infty$, and that $f'(x)$ and $f''(x)$ exist and are continuous over the whole range. Taking $y = x_{h-1}$ and $x = x_h$ in (15) and $y = x_h$ and $x = x_{h+1}$ in (16), we see that if we neglect terms of order $O(k^3)$, the equations

$$(x_h - x_{h-1})^2 P_h = (x_{h+1} - x_h)^2 P_{h+1}, \qquad h = 1, \cdots, (n-1),$$

that is,

$$(x_h - x_{h-1})^2 P_h = K_n, \qquad h = 1, \cdots, n, \tag{19}$$

where $K_n$ is a constant dependent on $f(x)$ and $n$, may be substituted for (18). This result may be compared with the approximate solution $(x_h - x_{h-1})P_h = C_n$ given in [2] to the similar problem of minimizing $\sum_1^n P_h\sigma_h$; we see by (11) that (19), and (11) of [2], may be replaced by $P_h\sigma_h^2 = K_n'$ and $P_h\sigma_h = C_n'$, respectively, without affecting the degree of approximation, whereby a certain analogy between the two results is discerned. Proceeding as in [2] we come to the same results as to the respective degree of approximation to the true minimizing values of the points satisfying (19) and the thereof resulting sum (17). We see that when $f(x_0) = 0$ or $x_0 = -\infty$ and/or $f(x_n) = 0$ or $x_n = +\infty$ we may substitute

$$8 f(x_1)\cdot(x_1 - \mu_1)^3 = K_n \quad \text{and/or} \quad = 8 f(x_{n-1})\cdot(\mu_n - x_{n-1})^3 \tag{20}$$

for those equations of (19) with $h = 1$ and/or $h = n$, also that $K_n$ varies with $n$ about as $n^{-3}$, and that an iterative method of finding $K_n$ may be employed. Finally we note that the methods of the last section of [2] may be used even in the present case, if we put $(x_h - \mu_h) = A_h$ and $(\mu_{h+1} - x_h) = B_{h+1}$, which results in

$$1 + \frac{\partial B_h}{\partial x_{h-1}} = -\frac{\partial A_h}{\partial x_{h-1}} = \frac{f(x_{h-1})\cdot(\mu_h - x_{h-1})}{P_h},$$

$$1 - \frac{\partial A_h}{\partial x_h} = \frac{\partial B_h}{\partial x_h} = \frac{f(x_h)\cdot(x_h - \mu_h)}{P_h}.$$

## REFERENCES

[1] Tore Dalenius, "The problem of optimum stratification," *Skand. Aktuarietids.*, Vol. 33 (1950), pp. 202–213.

[2] Gunnar Ekman, "An approximation useful in univariate stratification," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 219–229.

# ON THE MOMENTS OF THE TRACE OF A MATRIX AND APPROXIMATIONS TO ITS DISTRIBUTION

By K. C. S. Pillai[1] and Tito A. Mijares[2]

*The Statistical Center, University of the Philippines*

**1. Summary.** The first four moments of the sum of **s** non-null latent roots of a matrix occurring in multivariate analysis are studied. In particular, the first four moments of the sum of six roots are found, and are used to compare the upper percentage points obtained directly from the moment ratios with those from Pillai's approximate distribution.

**2. Introduction.** Distribution problems in multivariate analysis are often related to the distribution of the latent roots of a matrix derived from sample observations taken from multivariate normal populations. The form of the joint distribution of **s** non-null latent roots of a matrix in multivariate analysis as given by Roy [10], Hsu [3], and Fisher [2] is expressible as

$$f(\theta_1, \cdots, \theta_s) = C(s, m, n) \prod_{i=1}^{s} \theta_i^m (1 - \theta_i)^n \prod_{i>j} (\theta_i - \theta_j), \tag{2.1}$$

$$0 < \theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_s < 1,$$

where

$$C(s, m, n) = \frac{\pi^{s/2} \prod_{i=1}^{s} \Gamma\left(\frac{2m + 2n + s + i + 2}{2}\right)}{\prod_{i=1}^{s} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma\left(\frac{2n + i + 1}{2}\right) \Gamma\left(\frac{i}{2}\right)} \tag{2.2}$$

and **m** and **n** are defined differently for various situations described in [7].

Pillai [6], [8] has studied the distribution of the proposed test criterion, $V^{(s)} = \sum_{i=1}^{s} \theta_i$, deriving the first three moments and obtaining the fourth moment for $s = 2$, 3 and 4. He has also suggested [6], [8] an incomplete Beta distribution approximation to the distribution of $V^{(s)}$, and tabulated approximate percentage points for various values of the supplementary parameters **m** and **n** [9]. In this paper, the work of Pillai has been extended to generalize the fourth moment.

**3. The moment generating function of the sum of the roots.** The joint distribution given in (2.1) can be thrown into a determinantal form of the Vander-

Received August 5, 1958; revised January 19, 1959.

[1]United Nations Senior Adviser in Mathematical Statistics and Visiting Professor of Statistics. Now with the United Nations, New York.

[2]Associate Professor of Statistics.

monde type and the moment generating function for the sum of s non-null roots, $V^{(s)}$, can be expressed in the determinantal form [6], [8]

$$E(e^{tV^{(s)}})$$

$$(3.1) \qquad = C(s, m, n) \left| \begin{matrix} \int_0^1 \theta_s^{m+s-1}(1-\theta_s)^n e^{t\theta_s}\, d\theta_s \cdots \int_0^1 \theta_s^{m}(1-\theta_s)^n e^{t\theta_s}\, d\theta_s \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \int_0^{\theta_2} \theta_1^{m+s-1}(1-\theta_1)^n e^{t\theta_1}\, d\theta_1 \cdots \int_0^{\theta_2} \theta_1^{m}(1-\theta_1)^n e^{t\theta_1}\, d\theta_1 \end{matrix} \right| .$$

We may denote the pseudo-determinant (P.D.) [6], [8] in (3.1) by

$$U(m + s - 1, m + s - 2, \cdots, m; n; t)$$

and more conveniently, when $t = 0$, by $U(s - 1, s - 2, \cdots, 1, 0)$.

Differentiating (3.1) successively [1], [4] with respect to $t$ and setting $t = 0$ after each differentiation, we obtain

$$(3.2) \quad E(V^{(s)}) = \mu_1' = C(s, m, n)U(s, s-2, s-3, \cdots, 1, 0);$$

$$(3.3) \quad E[(V^{(s)})^2] = \mu_2' = C(s, m, n)[U(s+1, s-2, s-3, \cdots, 1, 0) + U(s, s-1, s-3, \cdots, 1, 0)];$$

$$(3.4) \quad \begin{aligned} E[(V^{(s)})^3] = \mu_3' = C(s, m, n)[&U(s+2, s-2, s-3, \cdots, 1, 0) \\ &+ 2\, U(s+1, s-1, s-3, \cdots, 1, 0) \\ &+ U(s, s-1, s-2, s-4, \cdots, 1, 0)]; \end{aligned}$$

$$(3.5) \quad \begin{aligned} E[(V^{(s)})^4] = \mu_4' = C(s, m, n)[&U(s+3, s-2, s-3, \cdots, 1, 0) \\ &+ 3\, U(s+2, s-1, s-3, \cdots, 1, 0) \\ &+ 2\, U(s+1, s, s-3, s-4, \cdots, 1, 0) \\ &+ 3\, U(s+1, s-1, s-2, s-4, \cdots, 1, 0) \\ &+ U(s, s-1, s-2, s-3, s-5, \cdots, 1, 0)]. \end{aligned}$$

The method can be extended for any number of differentiations. It may be noted that the first relation (3.2) for $\mu_1'$ contains only one P.D., since the other five vanish because two columns in each are equal.

**4. Values of the pseudo-determinants.** In this section we give the values of the P.D.'s involved in the expressions for the first four raw moments following (3.2)–(3.5). These were evaluated using a reduction formula by Pillai [8]. For details the reader is referred to [4].

Let us set $(2g + a)(2g + b) \cdots = G(a, b, \cdots)$ and $(m + n) = p$. Then the P.D. for the *first raw moment* is

$$(4.1) \qquad C(s, m, n)[U(s, s-2, s-3, \cdots, 1, 0)] = \frac{sM(s+1)}{P(2s+2)}.$$

For the *second raw moment*, we find

$$C(s, m, n)[U(s, s-1, s-3, s-4, \cdots, 1, 0)] = \frac{s(s-1)}{2!} \frac{M(s, s+1)}{P(2s+1, 2s+2)}, \tag{4.2}$$

$$C(s, m, n)[U(s+1, s-2, s-3, s-4, \cdots, 1, 0)] = \frac{s(s-1)}{2!} \frac{M(s, s+1)}{P(2s+1, 2s+4)} + \frac{sM(s+1, 2s+2)}{P(2s+2, 2s+4)}. \tag{4.3}$$

For the *third raw moment*, we find

$$C(s, m, n)[U(s, s-1, s-2, s-4, \cdots, 1, 0)] = \frac{s(s-1)(s-2)}{3!} \frac{M(s-1, s, s+1)}{P(2s, 2s+1, 2s+2)}, \tag{4.4}$$

$$\begin{aligned} C(s, m, n)&[U(s+1, s-1, s-3, \cdots, 1, 0)] \\ &= \frac{s(s-1)}{3!} \frac{M(s, s+1)}{P(2s, 2s+1, 2s+2, 2s+4)} \\ &\cdot \{4n[(2s-1)m + (s^2+2)] + 4(2s-1)m^2 + 12s^2m + (4s^3+2s^2+2s+4)\} \\ &+ \frac{m+s+1}{p+s+2} \cdot C(s, m, n)[U(s, s-1, s-3, s-4, \cdots, 1, 0)], \end{aligned} \tag{4.5}$$

$$\begin{aligned} C(s, m, n)&[U(s+2, s-2, s-3, \cdots, 1, 0)] \\ &= \frac{(s+1)s(s-1)}{3!} \frac{M(s, s+1, s+2)}{P(2s+1, 2s+2, 2s+6)} \\ &+ \frac{m+s+2}{p+s+3} \cdot C(s, m, n)[U(s+1, s-2, s-3, \cdots, 1, 0)] \end{aligned} \tag{4.6}$$

And for the *fourth raw moment*, we find

$$C(s, m, n)[U(s, s-1, s-2, s-3, s-5, \cdots, 1, 0)] = \frac{s(s-1)(s-2)(s-3)}{4!} \frac{M(s-2, s-1, s, s+1)}{P(2s-1, 2s, 2s+1, 2s+2)}, \tag{4.7}$$

$$\begin{aligned} C(s, m, n)&[U(s+1, s-1, s-2, s-4, s-5, \cdots, 1, 0)] \\ &= \frac{s(s-1)(s-2)}{3!2!} \frac{M(s-1, s, s+1)}{P(2s-1, 2s, 2s+1, 2s+2, 2s+4)} \\ &\cdot \{n[2(3s-1)m + (3s^2+s+10)] + 2(3s-1)m^2 \\ &+ (9s^2-s+2)m + (3s^3+s^2+2s+4)\} \\ &+ \frac{m+s+1}{p+s+2} \cdot C(s, m, n)[U(s, s-1, s-2, s-4, \cdots, 1, 0)], \end{aligned} \tag{4.8}$$

$$C(s, m, n)[U(s+1, s, s-3, s-4, \cdots, 1, 0)]$$

$$\begin{aligned} = \frac{s(s-1)}{3!2!} \frac{M(s, s+1)}{P(2s-1, 2s, 2s+1, 2s+2, 2s+3, 2s+4)} \\ \cdot \{n^2[16s(s+1)m^2 + 8s(2s^2+5s+9)m \\ + 4(s^4+4s^3+11s^2+8s+12)] \\ + n[32s(s+1)m^3 + 8s(8s^2+15s+13)m^2 \\ + 4s(10s^3+30s^2+45s+19)m \\ + 2(4s^5+17s^4+36s^3+31s^2+8s+12)] \\ + s(s+1)M(s+1, s+2, 2s, 2s+1)\}, \end{aligned} \tag{4.9}$$

$$C(s, m, n)[U(s+2, s-1, s-3, s-4, \cdots, 1, 0)]$$

$$\begin{aligned} = \frac{(s+1)s(s-1)}{3!2!} \frac{M(s, s+1, s+2)}{P(2s, 2s+1, 2s+2, 2s+3, 2s+6)} \\ \cdot \{n[2(3s-2)m + (3s^2-s+10)] + 2(3s-2)m^2 \\ + (9s^2+s-2)m + (3s^3+2s^2+s+6)\} \\ + \frac{m+s+2}{p+s+3} \cdot C(s, m, n)[U(s+1, s-1, s-3, \cdots, 1, 0)], \end{aligned} \tag{4.10}$$

$$C(s, m, n)[U(s+3, s-2, s-3, \cdots, 1, 0)]$$

$$\begin{aligned} = \frac{(s+2)(s+1)s(s-1)}{4!} \frac{M(s, s+1, s+2, s+3)}{P(2s+1, 2s+2, 2s+3, 2s+8)} \\ + \frac{m+s+3}{p+s+4} \cdot C(s, m, n)[U(s+2, s-2, \cdots, 1, 0)]. \end{aligned} \tag{4.11}$$

It may be noted that simplifications of relations (4.1) to (4.6) to obtain the first, second and third *central* moments yield exactly the results given by Pillai [6], [8]. The fourth central moment has not been obtained in general from (4.1) to (4.11); however, for particular cases, like that for $s = 6$ in the next section, these relations have been used to arrive at the expressions for $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$, where $\mu$'s denote central moments [4].

**5. Percentage points of $V^{(6)}$ using moment ratios and a Pearson curve approximation, and using Pillai's approximate beta distribution.** Putting $s = 6$ in relations (4.1) to (4.11) the following expressions for the $\beta$'s are obtained:

$$\beta_1 = \frac{4(n-m)^2(p+1)^2(p+8)(2p+13)}{3(2m+7)(2n+7)(p+4)(p+6)^2(p+9)^2}, \tag{5.1}$$

$$\begin{aligned} \beta_2 = \frac{(p+8)(2p+13)}{(2m+7)(2n+7)(p+4)^2(p+6)(p+9)(p+10)(2p+11)(2p+15)} \\ \cdot [4mn(6p^5 + 179p^4 + 2177p^3 + 13{,}176p^2 + 38{,}732p + 43{,}760) \\ + (92p^6 + 2996p^5 + 40{,}869p^4 + 294{,}677p^3 \\ + 1{,}169{,}444p^2 + 2{,}408{,}532p + 2{,}010{,}960)]. \end{aligned} \tag{5.2}$$

Table 5A shows the values of $\beta_1$ and $\beta_2$ for several values of **m** and **n**. The upper 5% points of $V^{(6)}$ given in Table 5B for each given **m** and **n** were obtained by interpolating in Table 42 of [5], "Percentage points of Pearson Curves for given $\beta_1$, $\beta_2$, expressed in standardized measure."

The percentage points from Pillai's approximate distribution [9] were obtained by referring to Snedecor's $F$ tables using the transformation

$$F = \frac{(2n + s + 1)}{(2m + s + 1)} \cdot \frac{V^{(s)}}{(s - V^{(s)})}, \tag{5.3}$$

with $f_1 = s(2m + s + 1)$ and $f_2 = s(2n + s + 1)$ degrees of freedom.

TABLE 5A

*Values of* $\begin{Bmatrix}\beta_1\\ \beta_2\end{Bmatrix}$ *of the exact distribution for* $s = 6$

| $n$ | $m$ | | | | |
|---|---|---|---|---|---|
| | 0 | 5 | 10 | 20 | 30 |
| 10 | 0.03919<br>3.00925 | 0.00381<br>2.97815 | 0.00000<br>2.97594 | 0.00424<br>2.98561 | 0.00990<br>2.99674 |
| 30 | 0.10177<br>3.11486 | 0.01615<br>3.01774 | 0.00990<br>2.99674 | 0.00213<br>2.98724 | 0.00000<br>2.98732 |
| 60 | 0.13690<br>3.18341 | 0.04460<br>3.05136 | 0.02192<br>3.01996 | 0.00725<br>3.00025 | 0.00260<br>2.99453 |
| 100 | 0.15553<br>3.21669 | 0.05814<br>3.07233 | 0.03013<br>3.03597 | 0.01269<br>3.01125 | 0.00636<br>3.00253 |

TABLE 5B

*Upper 5% points of* $V^{(6)}$ *using* (1) *the* $\beta$*'s of the exact distribution and a Pearson curve approximation, and* (2) *Pillai's approximation*

| $n$ | $m$ | | | | |
|---|---|---|---|---|---|
| | 0 | 5 | 10 | 20 | 30 |
| 10 (1) | 1.601 | 2.705 | 3.362 | 4.112 | 4.529 |
| (2) | 1.647 | 2.728 | 3.394 | 4.133 | 4.542 |
| 30 (1) | 0.762 | 1.460 | 1.983 | 2.729 | 3.239 |
| (2) | 0.771 | 1.469 | 2.002 | 2.747 | 3.247 |
| 60 (1) | 0.427 | 0.863 | 1.226 | 1.811 | 2.264 |
| (2) | 0.431 | 0.870 | 1.241 | 1.830 | 2.280 |
| 100 (1) | 0.269 | 0.559 | 0.813 | 1.249 | 1.615 |
| (2) | 0.271 | 0.563 | 0.822 | 1.259 | 1.628 |

It may be seen from Table 5B that the percentage points computed by the two methods practically agree in the first two places even for small values of $m$ and $n$. Further, we may also study the difference in probabilities corresponding to the percentage points from the two approximations. This may be done by considering the percentage points for approximation (1) in Table 5B and evaluating the probability in each case from Pillai's beta distribution approximation. For example, taking $m = 0$, $n = 10$ and percentage point 1.601, exact integration of the incomplete beta function gives the probability 0.9298 as against 0.95. However, for $m = 0$ and $n = 60$, with percentage point 0.427, the probability obtained by the same procedure is 0.9462. Hence, for larger values of $n$, as in the case of percentage points, the difference in probabilities also tends to be small.

**6. Acknowledgment.** The authors wish to acknowledge the facilities offered by the Statistical Center, University of the Philippines, for preparing this paper.

REFERENCES

[1] A. C. Aitken, *Determinants and Matrices*, Oliver & Boyd Ltd., Edinburgh, Great Britain, 1956.

[2] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.

[3] P. L. Hsu, "On the distribution of the roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.

[4] T. A. Mijares, "On the distribution of the sum of six roots of a matrix in multivariate analysis," unpublished M.S. thesis submitted in March, 1958 to the University of the Philippines Graduate School.

[5] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, Cambridge, 1954.

[6] K. C. S. Pillai, "On some distribution problems in multivariate analysis," Mimeograph Series No. 88, Institute of Statistics, University of North Carolina, 1954.

[7] K. C. S. Pillai, "Some new test criteria in multivariate analysis," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 117–121.

[8] K. C. S. Pillai, "Some results useful in multivariate analysis," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1106–1114.

[9] K. C. S. Pillai, *Concise Tables for Statisticians*, The Statistical Center, University of the Philippines, Manila, 1957.

[10] S. N. Roy, "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381–396.

# RANDOM GRAPHS

By E. N. Gilbert

*Bell Telephone Laboratories, Inc., Murray Hill, New Jersey*

**1. Introduction.** Let $N$ points, numbered $1, 2, \cdots, N$, be given. There are $N(N-1)/2$ lines which can be drawn joining pairs of these points. Choosing a subset of these lines to draw, one obtains a graph; there are $2^{N(N-1)/2}$ possible graphs in total. Pick one of these graphs by the following random process. For all pairs of points make random choices, independent of each other, whether or not to join the points of the pair by a line. Let the common probability of joining be $p$. Equivalently, one may erase lines, with common probability $q = 1 - p$ from the complete graph.

In the random graph so constructed one says that *point $i$ is connected to point $j$* if some of the lines of the graph form a path from $i$ to $j$. If $i$ is connected to $j$ for every pair $i, j$, then the graph is said to be *connected*. The probability $P_N$ that the graph is connected, and also the probability $R_N$ that two specific points, say 1 and 2, are connected, will both be found.

As an application, imagine the $N$ points to be $N$ telephone central offices and suppose that each pair of offices has the same probability $p$ that there is an idle direct line between them. Suppose further that a new call between two offices can be routed via other offices if necessary. Then $R_N$ is the probability that there is some way of routing a new call from office 1 to office 2 and $P_N$ is the probability that each office can call every other office.

Exact expressions for $P_N$ and $R_N$ are given in Section 2. These results are unwieldy for large $N$. Bounds on $P_N$ and $R_N$ derived in Section 3 show that

$$P_N \sim 1 - Nq^{N-1} \tag{1}$$

and

$$R_N \sim 1 - 2q^{N-1} \tag{2}$$

asymptotically as $N \to \infty$.

Other related results appear in a paper by Austin, Fagen, Penney, and Riordan [1]. These authors use a different random process to pick a graph and they find a generating function for the distribution of the number of connected pieces in the random graph.

**2. Exact results.** $P_N$ may be expressed in terms of the number $C_{N,L}$ of connected graphs having $N$ labeled points and $L$ lines. Since each such graph has probability $p^L q^{-L+N(N-1)/2}$ of being the chosen graph, it follows that

$$P_N = \sum_L C_{N,L} p^L q^{-L+N(N-1)/2}.$$

Received January 26, 1959; revised July 29, 1959.

In this formula the range of summation is $N - 1 \leqq L \leqq N(N-1)/2$. In [3] and [4] a generating series for $C_{N,L}$ was given in the form

$$\sum_{N,L}^{\infty} C_{N,L} \frac{x^N y^L}{N!} = \log\left(1 + \sum_{i=1}^{\infty} \frac{x^i(1+y)^{i(i-1)/2}}{i!}\right).$$

This result is easily converted into a generating series for $P_N$, viz.,

$$\sum_{N=1}^{\infty} P_N \frac{x^N q^{-N(N-1)/2}}{N!} = \log\left(1 + \sum_{i=1}^{\infty} \frac{x^i q^{-i(i-1)/2}}{i!}\right). \tag{3}$$

It may be noted that, when $0 \leqq q < 1$ and $x \neq 0$, neither series in (3) converges. The equality in (3) merely signifies that $P_N$ may be found by formally expanding the logarithm into a power series and collecting coefficients of $x^N$. One can perform the expansion analytically to obtain an explicit formula

$$P_N = \sum_{r_1,\cdots,r_N} \frac{(-1)^{n-1}(n-1)!N!q^{(N^2-1^2r_1-\cdots-N^2r_N)/2}}{r_1!\cdots r_N!(1!)^{r_1}\cdots(N!)^{r_N}}.$$

The sum is extended over all non-negative integer solutions of $r_1 + 2r_2 + \cdots + Nr_N = N$ (i.e. over all partitions of $N$). The letter $n$ in the sum is $n = r_1 + \cdots + r_N$.

The first few instances of this formula are

$$P_1 = 1$$

$$P_2 = 1 - q$$

$$P_3 = 1 - 3q^2 + 2q^3$$

$$P_4 = 1 - 4q^3 - 3q^4 + 12q^5 - 6q^6$$

$$P_5 = 1 - 5q^4 - 10q^6 + 20q^7 + 30q^8 - 60q^9 + 24q^{10}$$

$$P_6 = 1 - 6q^5 - 15q^8 + 20q^9 + 120q^{11} - 90q^{12} - 270q^{13} + 360q^{14} - 120q^{15}.$$

For larger values of $N$ the number of terms in the formula for $P_N$ increases rapidly. $P_N$ may then be computed more easily by means of the recurrence relation

$$1 - P_N = \sum_{k=1}^{N-1} \binom{N-1}{k-1} P_k q^{k(N-k)}. \tag{4}$$

The $k$th term of (4) is the probability that point 1 is connected to exactly $k - 1$ of the $N - 1$ other points. Then (4) follows by noting that point 1 is connected to 0, 1, $\cdots$, or $N - 1$ other points with probability 1.

The argument which was used to derive (4) may be modified to give the following formula for $R_N$:

$$1 - R_N = \sum_{k=1}^{N-1} \binom{N-2}{k-1} P_k q^{k(N-k)}. \tag{5}$$

TABLE 1

| $q =$ | .1 | .3 | .5 | .7 | .9 |
|---|---|---|---|---|---|
| $P_2$ | .90000 | .70000 | .50000 | .30000 | .10000 |
| $P_3$ | .97200 | .78400 | .50000 | .21600 | .02800 |
| $P_4$ | .99581 | .89249 | .59375 | .21865 | .01293 |
| $P_5$ | .99949 | .95751 | .71094 | .25626 | .00810 |
| $P_6$ | .99994 | .98497 | .81569 | .31690 | .00624 |
| $R_2$ | .90000 | .70000 | .50000 | .30000 | .10000 |
| $R_3$ | .98100 | .84700 | .75000 | .36300 | .10900 |
| $R_5$ | .99980 | .98143 | .85353 | .52528 | .13134 |
| $R_7$ | .9999980 | .99850 | .96302 | .70634 | .16118 |

The $k$th term of (5) is the probability that point 1 is connected to exactly $k$ of the $N - 2$ points $3, \cdots, N$. Then the sum is the probability that points 1 and 2 are not connected.

Using these results, R. W. Hamming and the author computed numerical values of $P_N$ and $R_N$ which appear in Table 1.

**3. Bounds.** The formulas of Section 2 solve the problem for small $N$ only. In this section we estimate $P_N$ and $R_N$ for large $N$. As $N$ increases, the number of paths by which points 1 and 2 may be joined increases. Then it is not surprising that $R_N \to 1$ as $N \to \infty$ for every fixed $p > 0$. That $P_N \to 1$ too is less obvious since increasing $N$ also increases the number of pairs of points to be connected. Indeed, Table 1 shows $P_N$ decreasing for $N \leqq 6$ when $q = .9$. The more precise results (1) and (2) follow from the bounds which we now derive.

THEOREM 1:

$$\left\{1 - \frac{N-1}{2} q^{N-1}\right\} N q^{N-1} \leqq 1 - P_N$$

*and*

$$1 - P_N \leqq q^{N-1}\{(1 + q^{(N-2)/2})^{N-1} - q^{(N-2)(N-1)/2}\} + q^{N/2}\{(1 + q^{(N-2)/2})^{N-1} - 1\}.$$

THEOREM 2:

$$(2 - q^{N-2})q^{N-1} \leqq 1 - R_N \leqq 2\, q^{N-1}(1 + q^{(N-2)/2})^{N-2}.$$

The lower bound in Theorem 2 is just the probability that at least one of the two points 1, 2 is connected to no other point.

A similar idea is used in Theorem 1. A lower bound on $1 - P_N$ is the probability $T$ that at least one of the points $1, 2, \cdots, N$ is connected to no other point. Let $E_i$ denote the event that point $i$ is connected to no other point; then $T$ is the union of the events $E_1, \cdots, E_N$. A lower bound on $T$ (and hence on

$1 - P_N$) is provided by an inequality of Bonferroni (see Feller [2], p. 100):

$$\sum_i P(E_i) - \sum_{i<j} P(E_iE_j) \leqq T.$$

Since $P(E_i) = q^{N-1}$ and Pr $(E_iE_j) - q^{2N-3}$, we obtain the lower bound stated.

The upper bounds are obtained using (4) and (5). In both cases we bound $P_k$ by 1. To bound $q^{k(N-k)}$ we use the fact that $x(N - x)$ is a convex function of $x$. Then

$$k(N-k) \geqq \frac{(N-2)k + N}{2} \qquad \text{if } 1 \leqq k \leqq \frac{N}{2},$$

$$k(N-k) \geqq \frac{(N-2)(N-k) + N}{2} \qquad \text{if } \frac{N}{2} \leqq k \leqq N-1,$$

and

$$q^{k(N-k)} \leqq q^{N/2}\{q^{(N-2)k/2} + q^{(N-2)(N-k)/2}\}$$

for $1 \leqq k \leqq N - 1$. When these bounds are inserted into (4) and (5), the sums reduce to the expression shown in Theorems 1 and 2.

When $N$ becomes large the bounds are in close agreement. It follows from Theorems 1 and 2 that

$$P_N = 1 - Nq^{N-1} + O(N^2q^{3N/2}),$$

and

$$R_N = 1 - 2q^{N-1} + O(Nq^{3N/2}).$$

Checking these approximate formulas against $P_N$ and $R_N$ in Table 1, it appears likely that $Nq^{N-1}$ and $2q^{N-1}$ will represent $1 - P_N$ and $1 - R_N$ to within 3% when $q \leqq .3$ and $N \geqq 6$. For the same degree of approximation, larger values of $q$ will require larger values of $N$.

## REFERENCES


[1] T. L. Austin, R. E. Fagen, W. F. Penny, and John Riordan, "The number of components in random linear graphs," *Ann. Math. Stat.*, Vol. 30 (1959) pp. 747–754.

[2] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, Second Edition, Wiley, New York, 1957.

[3] E. N. Gilbert, "Enumeration of labelled graphs," *Canadian J. Math.*, Vol. 8 (1956), pp. 405–411.

[4] R. J. Riddell, Jr. and G. E. Uhlenbeck, "On the virial development of the equation of state of monoatomic gases," *J. Chem. Phys.*, Vol. 21 (1953), pp. 2056–2064.

# SCALE MIXING OF SYMMETRIC DISTRIBUTIONS WITH ZERO MEANS[1]

By E. M. L. Beale[2] and C. L. Mallows[3]

*Princeton University*

**0. Summary.** Suppose that a distribution $A$ is a mixture of distributions similar to $B$ but with different scale parameters; or (almost equivalently) that a distribution $F$ is a convolution of a given distribution $G$ with some other distribution. We derive conditions on (i) the moments of $A$ and $F$ and (ii) on the derivatives of $A$ and $F$; these conditions are necessary, but are not sufficient in general. The conditions (ii) are appropriate when $B$ (or $G$) is of Pólya type 3.

**1. Introduction.** Suppose $A(x)$ and $B(x)$ are cumulative distribution functions (c.d.f.'s) on the real line, continuous on the right, and a.e. symmetric about the origin, so that

$$A(x) + A(-x - 0) = 1 = B(x) + B(-x - 0), \qquad -\infty < x < \infty. \tag{1}$$

We write $X_A$ for a random variable (r.v.) having the c.d.f. $A(x)$, and similarly for $X_B$. We shall say that $A(x)$ is a $B$-mixture if there exist r.v.'s $X_A$, $X_B$, and $Y$, where $Y$ is non-negative and independent of $X_B$, such that

$$X_A = X_B Y \tag{2}$$

or equivalently, if there exists a c.d.f. $C(\sigma^2)$ on $[0, \infty)$, continuous on the right, such that

$$A(x) = \int_0^\infty B(x/\sigma)\, dC(\sigma^2), \qquad 0 < x \tag{3}$$

where we interpret $B(x/0)$ as 1 for $x > 0$. It is clear that $A(x)$ is discontinuous at $x = 0$ if $C(0) \neq 0$.

In a closely related situation (see Section 3 below), if $F(x)$ and $G(x)$ are c.d.f.'s on the real line (not necessarily symmetric), we shall say that $F$ is a $G$-convolution if there exist r.v.'s $X_F$, $X_G$, and $Z$ ($Z$ having c.d.f. $H(x)$ and being independent of $X_G$, not necessarily non-negative) such that

$$X_F = X_G + Z. \tag{4}$$

Some general theorems concerning the existence and measurability of functions related to mixtures of distributions were proved by Robbins [3]. Teichroew [6] considered the case where $B(x)$ is the unit Normal c.d.f., and $C(\sigma^2)$ is of Pearson type III.

---

Received September 15, 1958; revised March 10, 1959.

[1] Research at Princeton University sponsored by the Office of Ordnance Research, U. S. Army; Statistical Techniques Research Group, contract No. DA-36-034-ORD 2297.

[2] Now at Admiralty Research Laboratory, Teddington, Middlesex, England.

[3] Now at University College London, England.

Our interest in the mixture problem arose out of some research where it was possible to prove that a certain procedure was optimal whenever an error distribution was Normal with zero mean but arbitrary variance, and also whenever it was a mixture of such distributions. It was thus of interest to determine as far as possible the properties of such mixtures. In general, given $A$ and $B$, we would like to be able to determine whether or not $A$ can be regarded as a $B$-mixture; and similarly for the convolution problem.

Hirschman and Widder [1] investigate (4) at great length, but their results are not in the form we desire; thus for the case where $X_G$ is normal (with mean zero and variance $v$, say) they give two sets of necessary and sufficient conditions for (4) to hold. The first of these ([1] Theorem 12.2) requires a knowledge of $\partial F(x)/\partial v$, and the second ([1] Theorem 12.4) achieves the inversion of (4) by means of an infinite series of derivatives of $F(x)$; the required condition is that the sum of this series be everywhere non-decreasing (i.e. gives a c.d.f.). This last formula has been much used in practice; see e.g. Smart [5].

We assume that the distribution $A$ (or $F$) is completely known; we do not say anything about the statistical problems of testing whether a random sample can reasonably be assumed to come from some B-mixture, and if so of estimating the mixing distribution $C$. Robbins [4] considers this estimation problem. He remarks that it is of considerable importance in other connections, but awaits a satisfactory practical solution.

In Section 2 we derive necessary and sufficient conditions for the existence of some $A$ that is a $B$-mixture having given moments through order $2r$. In Section 3 we examine the relation between the mixture problem and the convolution problem. In Section 4 we obtain a necessary condition for a given $A$ to be a $B$-mixture (or for a given $F$ to be a $G$-convolution) in terms of the frequency functions $A'(x)$ (or $F'(x)$) and their derivatives; the validity of these conditions depends on certain properties of the derivatives of $B$ (or $G$), related to the theory of Pólya types; this relation is explored in Section 5.

**2. Conditions on moments.** From (2) we have immediately that if $A$ is a $B$-mixture, then

$$E(X_A^{2r}) = E(X_B^{2r})E(Y^{2r}) \tag{5}$$

and the l.h.s. exists if and only if each of the factors on the r.h.s. is finite. Since $Y^2$ is to be a r.v. on $[0, \infty]$, its moments must satisfy certain inequalities, the simplest of which is the obvious one $E(Y^4) \geqq \{E(Y^2)\}^2$. Hence we obtain necessary relations between the moments of $A$ and $B$; the simplest is

$$\mu_4(A)/\mu_2(A)^2 \geqq \mu_4(B)/\mu_2(B)^2. \tag{6}$$

so that the kurtosis of a mixture is never less than the kurtosis of a single component. Conversely, these relations are sufficient for the existence of some distribution $A(x)$ that is a $B$-mixture, having the given moments.

The result that a mixture of Normal distributions (with zero means) is necessarily leptokurtic (unless it reduces to a single Normal distribution) seems to be

widely known, though apparently unpublished. It is worth bearing in mind when considering the argument that practical error distributions "must" tend to Normality because the error is the sum of many independent components. It is arguable that many error distributions are mixtures of distributions with a common mean but different variances, and can therefore be expected to be leptokurtic.

**3. The distribution of** $\ln |X_A|$. Another simple line of approach to the mixture problem is to consider the distribution of $\ln |X_A|$. Before we can do this we must consider the probability that $X_A = 0$, since $\ln X_A$ is then undefined. Writing $A_0 = \Pr\{X_A \neq 0\}$ and similarly for $B_0$ and $C_0$, we have immediately from (2) that

$$A_0 = B_0 C_0 . \tag{7}$$

Also from (2), conditioned that none of the r.v.'s are zero (i.e. $X_A \neq 0$), we have

$$\ln |X_A| = \ln |X_B| + \ln Y \tag{8}$$

which is exactly (4). Thus we have transformed the mixture problem into the convolution problem. If we define the conditional characteristic functions of $\ln |X_A|$ and $\ln |X_B|$ by

$$\varphi_A(t) = \frac{2}{A_0}\int_{0+}^{\infty} x^{it}\, dA(x), \qquad \varphi_B(t) = \frac{2}{B_0}\int_{0+}^{\infty} x^{it}\, dB(x), \tag{9}$$

then we have from (7) and (8)

THEOREM 1: *A necessary and sufficient condition for A to be a B-mixture is that* $A_0 \leqq B_0$ *and* $\varphi_A(t)/\varphi_B(t)$ *is the ch.fn. of some distribution on* $(-\infty, \infty)$.

In a sense this is the complete answer to the problem, but unfortunately the criterion is not in general easy to apply. In some circumstances a numerical approach based on (8) may be effective. An approach via the moments of $\ln |X_A|$ and $\ln |X_B|$ (similar to that in the previous section) will yield a series of necessary conditions.

**4. Conditions on the frequency function.** We now consider criteria based on derivatives of the c.d.f.'s. Let us assume that $B(x)$ is four times differentiable everywhere, and that $b = B'(x) > 0$ for all $x$. It will follow that any $B$-mixture $A(x)$ is four times differentiable everywhere except perhaps at $x = 0$, and that $a(x) = A'(x) > 0$ wherever this exists.

Now $A$ is assumed to be a mixture of distributions with zero mean and varying scale parameter $\sigma$; so that part of the distribution $A$ near $x = 0$ will consist primarily of those components with small $\sigma$, while the part with $|x|$ large will consist primarily of components with large $\sigma$. We may expect to find a necessary condition for $A$ to be a $B$-mixture based on this fact, and the simplest such condition seems to be the following:

*Conjecture*: If one assumed that only one component contributed to $a(x)$ for

a particular $x$, and one estimated the value of $\sigma$ for this component from the values of $a(x)$ and $a'(x)$, then for any $A$ that is a $B$-mixture, the value of $\sigma$ so defined is a non-decreasing function of $|x|$.

The value of $\sigma$ described is defined by the equation

$$(x/\sigma)b'(x/\sigma)/b(x/\sigma) = xa'(x)/a(x). \tag{10}$$

If this equation holds for more than one $\sigma$, we could make the estimate unique by agreeing to take the smallest value satisfying (10). But we can hardly expect the conjecture to be true unless $xb'(x)/b(x)$ is a strictly monotone function of $x$. This is equivalent to the condition that the distribution $B$ is strictly of Pólya type 2 (monotone likelihood ratio) with respect to the parameter $\sigma$, as defined by Karlin [2].

It turns out (see Section 5) that the conjecture is correct if $B$ is also of Pólya type 3 with respect to $\sigma$. Although it is possible to construct symmetrical distributions that are not Pólya type 3 with respect to $\sigma$, almost all the principal cases occurring in statistical practice—such as the Normal, double-exponential, Cauchy, rectangular, triangular—are of this type.

In terms of the distribution of $\ln |X_A|$ and $\ln |X_B|$, the conjecture asserts that if $F$ is a $G$-convolution, and writing $f(x) = F'(x)$, $g(x) = G'(x)$, then the value of $\mu$ defined by

$$g'(x-\mu)/g(x-\mu) = f'(x)/f(x) \tag{11}$$

is a non-decreasing function of $x$. In the following, we shall work in terms of the convolution problem. We shall write

$$R_j(x) = g^{(j)}(x)/g(x) \tag{12}$$

so that

$$R_1' = dR_1/dx = R_2 - R_1^2. \tag{13}$$

THEOREM 2: *If for all* $x$, (i) $g(x) > 0$, (ii) $dR_1/dx < 0$, (iii) $R_2(x)$ *is a convex function of* $R_1(x)$, *then* $\mu$, *defined by* (11), *is a non-decreasing function of* $x$.

*Conversely, given* (i) *and* (ii), *if* $\mu$ *is non-decreasing for all G-convolutions, then* (iii) *must hold.*

The statement of the theorem for the mixture problem, with $\sigma$ defined by (10), is the same as this with the $R$'s defined as

$$R_1(x) = 1 + x\frac{b'(x)}{b(x)}, \qquad R_2(x) = 1 + 3x\frac{b'(x)}{b(x)} + x^2\frac{b''(x)}{b(x)}. \tag{14}$$

PROOF: From (11),

$$R_1(x-\mu) = f'(x)/f(x) \tag{15}$$

$$= \frac{\int R_1(x-m)g(x-m)\,dH(m)}{\int g(x-m)\,dH(m)}, \tag{16}$$

which shows that $\mu$ exists (by (ii)). Multiplying (15) by $f(x)$ and differentiating with respect to $x$, we find

$$f'(x)R_1(x-\mu) + f(x)R_1'(x-\mu)(1 - d\mu/dx) = f''(x)$$

so that (using (13) and (15))

$$\begin{aligned} -f(x)R_1'(x-\mu)\,d\mu/dx &= f''(x) - f(x)R_2(x-\mu) \\ &= \int \{R_2(x-m) - R_2(x-\mu)\}g(x-m)\,dH(m). \end{aligned} \tag{17}$$

Now $f(x) > 0$, and $R_1'(x-\mu) < 0$ by (ii); further, (iii) implies that for each $x - \mu$ there exists some number $k$ (independent of $m$) such that for all $m$

$$R_2(x-m) - R_2(x-\mu) \geqq k\{R_1(x-m) - R_1(x-\mu)\}. \tag{18}$$

But from (16)

$$\int \{R_1(x-m) - R_1(x-\mu)\}g(x-m)\,dH(m) = 0 \tag{19}$$

so that the r.h.s. of (17) is $\geqq 0$, and $d\mu/dx \geqq 0$ as required.

Conversely, suppose (iii) is false. We shall construct a $G$-convolution which has $d\mu/dx < 0$ at $x = 0$. By our assumption, there exist $m_1$, $m_2$, $\mu$ (with $m_1 < \mu < m_2$) such that

$$\tfrac{1}{2}\{R_1(-m_1) + R_1(-m_2)\} = R_1(-\mu), \tag{20}$$

$$\tfrac{1}{2}\{R_2(-m_1) + R_2(-m_2)\} < R_2(-\mu). \tag{21}$$

Now choose $H(m)$ so that $dH(m)/dm = 0$ except at $m = m_1$ and $m_2$, with

$$dH(m_i) = \frac{1}{g(-m_i)}\left\{\frac{1}{g(-m_1)} + \frac{1}{g(-m_2)}\right\}^{-1}, \qquad i = 1, 2. \tag{22}$$

Then by (20), (16) is satisfied (for $x = 0$), and by (21), the r.h.s. of (17) is $<0$, so that $d\mu/dx < 0$.

Theorem 2 provides us with a necessary condition for $F$ to be a $G$-convolution (or for $A$ to be a $B$-mixture); namely, the $\mu$ (or $\sigma$) defined by (11) (or (10)) must be a non-decreasing function of $x$. Unfortunately it will not provide a sufficient condition unless $R_2$ is a linear function of $R_1$; and this is impossible over the whole range of $x$. ($R_2$ can be a piecewise linear function of $R_1$ if we allow $d^3g/dx^3$ to be discontinuous.) However, relaxing condition (i) of the theorem, we can obtain distributions for which $R_2$ is a linear function of $R_1$ wherever $g(x) > 0$; two such distributions for the mixtures problem are the rectangular and the triangular. It is easy to verify that a necessary and sufficient condition for $A$ to be a mixture of rectangular distributions is that $A$ be unimodal; and that necessary and sufficient conditions for $A$ to be a mixture of triangular distributions are that $A$ should have a derivative $a(x)$ everywhere except possibly at $x = 0$, while $a_+'(x)$ exists and is non-positive and non-decreasing for all $x > 0$. If $b(x)$ (or $g(x)$) $> 0$ for all $x$, then, for example, no distribution $A$ for which

the estimate of $\sigma$ is constant for all $x > x_0$, but takes a different value for some smaller value of $x$, can be a $B$-mixture.

If $B$ is a Normal distribution (for which the conditions of Theorem 2 are satisfied), the result can be utilized in the form of a log/square plot, in which $\log a(x)$ is plotted as a function of $x^2$. It is easy to see that the slope of this curve is inversely proportional to the estimate of $\sigma$; so the theorem shows that a necessary condition for the given distribution $A(x)$ to be a mixture of Normal distributions is that the log/square plot be convex.

**5. Relation to theory of Pólya types.** The conditions required in Theorem 2 can be expressed in terms of the determinants

$$\Delta_n = \left| \frac{\partial^i}{\partial x^i} \frac{\partial^j}{\partial \mu^j} p(x, \mu) \right|_{i,j=0}^{n-1} \tag{23}$$

for $n = 1, 2, 3$ with $p(x, \mu) = g(x - \mu)$. We shall prove

THEOREM 3: *The conditions* (i), (ii), (iii) *of Theorem* 2 *are equivalent to*

$$\Delta_1 > 0, \qquad \Delta_2 > 0, \qquad \Delta_3 \geqq 0.$$

PROOF: It is easy to see that the signs of these determinants are unaffected by a monotonic increasing transformation of either the independent variable $x$ or the parameter $\mu$; so that a proof of the theorem for the convolution problem will imply the corresponding result for the mixtures problem also. In the following, the argument of all the functions involved is $x - \mu$.

For $n = 1$, (23) gives $g > 0$, which is (i). For $n = 2$, (23) gives

$$\begin{vmatrix} g & -g' \\ g' & -g'' \end{vmatrix} > 0, \quad \text{i.e.} \quad \begin{vmatrix} 1 & R_1 \\ R_1 & R_2 \end{vmatrix} < 0, \tag{24}$$

i.e. by (13), $R_1' < 0$, which is (ii). Now

$$d^2R_2/dR_1^2 = (R_1')^{-3}(R_1'R_2'' - R_1''R_2') \tag{25}$$

so that condition (iii) is equivalent to

$$\begin{vmatrix} 1 & 0 & 0 \\ R_1 & R_1' & R_1'' \\ R_2 & R_2' & R_2'' \end{vmatrix} \leqq 0 \tag{26}$$

By differentiation we have successively

$$\begin{aligned} g' &= gR_1, \quad & g'' &= g'R_1 + gR_1', \quad & g''' &= g''R_1 + 2g'R_1' + gR_1'', \\ g'' &= gR_2, \quad & g''' &= g'R_2 + gR_2', \quad & g'''' &= g''R_2 + 2g'R_2' + gR_2''. \end{aligned} \tag{27}$$

Hence manipulating the determinant in (26) according to the scheme

$$\begin{aligned} (\text{col } 3)' &= g(\text{col } 3) + 2g'(\text{col } 2) + g''(\text{col } 1) \\ (\text{col } 2)' &= g(\text{col } 2) + g'(\text{col } 1) \end{aligned} \tag{28}$$

we obtain

$$\begin{vmatrix} g & g' & g'' \\ g' & g'' & g''' \\ g'' & g''' & g'''' \end{vmatrix} \leqq 0 \tag{29}$$

which is equivalent to $\Delta_3 \geqq 0$.

Karlin [2] calls a family of distributions

$$P(x, \mu) = \beta(\mu) \int_{-\infty}^{x} p(x, \mu)\, d\lambda(x) \tag{30}$$

of Pólya type $m$ (strictly of Pólya type $m$) if the determinants

$$D_n = | p(x_i, \mu_j) |_{i,j=1}^{n} \tag{31}$$

are $\geqq 0$ ($> 0$) for $n = 1, 2, \cdots, m$, for all

$$x_1 < x_2 < \cdots < x_n, \qquad \mu_1 < \mu_2 < \cdots < \mu_n. \tag{32}$$

Karlin shows that Pólya $m$ implies $\Delta_n \geqq 0$ ($n = 1, \cdots, m$), while $\Delta_n > 0$ ($n = 1, \cdots, m$) implies strict Pólya $m$.

We are indebted to the referee for the following remarks. One can derive only $\Delta_n \geqq 0$ when assuming strict Pólya $m$, with $\Delta_n > 0$ for almost all $x$ and $\mu$. It is true however that if $p(x, \mu) = p(x - \mu)$ (as is the case in the present problem), then the equivalence is correct. This last result is quite deep and is not published in the literature. Most strict Pólya type distributions satisfy $\Delta_n > 0$ everywhere, but there may be isolated points where equality takes place.

Thus our conditions (i) and (ii) are equivalent to strict Pólya 2, and (iii) is implied by Pólya 3.

Karlin [2] remarks that if $\Delta_n \geqq 0$ ($n = 1, \cdots, m$) with strict inequality almost everywhere, then under a certain weak assumption the convolution of $G(x)$ with a Normal distribution of arbitrarily small variance $\sigma^2$ will be strictly Pólya $m$, and hence (taking the limit as $\sigma^2$ tends to zero) $G$ will be Pólya $m$. In such cases Theorem 2 can still be applied, provided that, whenever (11) does not define $\mu$ uniquely, $\mu$ is taken as the appropriate limit as $\sigma^2$ tends to zero.

The authors are grateful to the referee for his suggestions for improving the presentation of the paper, and for clarifying the situation in the last section.

## REFERENCES

[1] I. I. Hirschman and D. V. Widder, *The Convolution Transform*, Princeton University Press, Princeton, 1955.

[2] Samuel Karlin, "Pólya type distributions, II", *Ann. Math. Stat.* Vol. 28 (1957), pp. 281–308.

[3] Herbert Robbins, "Mixture of distributions," *Ann. Math. Stat.* Vol. 19 (1948), pp. 360–369.

[4] Herbert Robbins, "An empirical Bayes approach to statistics," *Third Berkeley Symposium on Mathematical Statistics and Probability* Vol. 1 (1956), pp. 157–163.

[5] W. M. Smart, *Stellar Dynamics*, Cambridge University Press, 1938.

[6] D. Teichroew, "The mixture of normal distributions with different variances," *Ann. Math. Stat.* Vol. 28 (1957), pp. 510–512.

## MEASURABILITY OF EXTENSIONS OF CONTINUOUS RANDOM TRANSFORMS

By Otto Hanš

*Czechoslovak Academy of Sciences, Prague*

**1. Summary.** Extension theorems of Tietze and Hahn-Banach play an important part in functional analysis. It seems reasonable to deal with similar questions for random transforms. In the present paper some measurability problems arising in connection with this probabilistic generalization are solved.

**2. Introduction.** First of all we shall introduce some convenient notions, definitions of which follow those given in [1].

Let $(\Omega, \mathfrak{S})$ and $(Z, \mathfrak{Z})$ be two measurable spaces and $U$ a mapping of the space $\Omega$ into the space $Z$ so that the inclusion

$$\{\{\omega: U(\omega) \varepsilon B\}: B \varepsilon \mathfrak{Z}\} \subset \mathfrak{S}$$

holds. Then the mapping $U$ will be called a generalized random variable, or, more precisely, a generalized random variable with values in the space $(Z, \mathfrak{Z})$.

If $(\Omega, \mathfrak{S})$ and $(Z, \mathfrak{Z})$ are two measurable spaces, $X$ an arbitrary non-empty set and $T$ a mapping of the Cartesian product $\Omega \times X$ into the space $Z$ satisfying the condition

$$\{\{\omega: T(\omega, x) \varepsilon B\}: x \varepsilon X, B \varepsilon \mathfrak{Z}\} \subset \mathfrak{S},$$

then we shall speak about a random transform, or, more precisely, about a random transform of the Cartesian product $\Omega \times X$ into the space $(Z, \mathfrak{Z})$.

Let us remark that in case $Z$ is a metric space, we usually choose the $\sigma$-algebra $\mathfrak{Z}$ as the class of all Borel subsets of the space $Z$. Under this additional agreement about the $\sigma$-algebra $\mathfrak{Z}$, a number of theorems and criteria have been stated in [1]. For the purposes of the present paper Criterion 6 is of most importance:

If $Z$ is a separable Banach space then a mapping $U$ is a generalized random variable if, and only if, for every bounded linear functional $f$ from a subset $\Delta$ of the first adjoint Banach space $Z^*$, where the subset $\Delta$ is total on the whole Banach space $Z$, the compound mapping $f(U)$ is a real-valued random variable.

Some other definitions of a generalized random variable (or of a random element) have been given by other authors. Thus, for instance, Mourier [2] defines a random element only in the case $Z$ is a Banach space in the following way: a mapping $U$ is a random element if for every bounded linear functional $f$ from the first adjoint space $Z^*$ the compound mapping $f(U)$ is a real-valued random variable. Though for separable Banach spaces the definition of Mourier and the one of ours coincide, for arbitrary Banach spaces they differ. The definition of Mourier enables one to prove that the sum of random elements is again

---

Received January 14, 1958; revised April 3, 1958.

a random element, while for generalized random variables this statement need not hold as shown by Nedoma in [3]. On the other hand generalized random variables possess the important property that each compound mapping $\tau(U)$ formed by means of a Borel measurable mapping $\tau$ (of the measurable space $(Z, \mathfrak{Z})$ into another measurable space $(Y, \mathfrak{Y})$) and a generalized random variable $U$ is a generalized random variable (with values in the measurable space $(Y, \mathfrak{Y})$).

Bharucha-Reid [4] follows essentially the definition of Mourier, provided his random elements have values in Orlicz spaces.

The conception of Kolmogoroff and Prochorow [5] is a generalization of the notion of a stochastic process, while Dubins [6] defines a generalized random variable as a homomorphism of some Boolean algebra into the measure ring induced by some probability space.

Let us remark that our definition does not depend on any probability measure defined on the measurable space $(\Omega, \mathfrak{S})$ and this is sometimes an advantage.

**3. Probabilistic Tietze theorem.** In what follows $R$ denotes the space of all real numbers and $\mathfrak{R}$ the $\sigma$-algebra of all Borel subsets of the space $R$.

THEOREM 1: *Let $(\Omega, \mathfrak{S})$ be a measurable space, $X$ a separable metric space, $M$ a closed subset of the space $X$ and $V$ a random transform of the Cartesian product $\Omega \times X$ into the space $(R, \mathfrak{R})$, which is for every fixed $\omega \in \Omega$ a continuous mapping $V(\omega, \cdot)$ of the set $M$ into the space $R$, such that for every couple $(\omega, x) \in \Omega \times M$ the relation $|V(\omega, x)| \leqq s(\omega)$, where $s$ is a real-valued random variable, holds.*

*Then there exists a random transform $T$ of the Cartesian product $\Omega \times X$ into the space $(R, \mathfrak{R})$ so that*

(i) *for every couple $(\omega, x) \in \Omega \times M$ we have $T(\omega, x) = V(\omega, x)$;*

(ii) *for every $\omega \in \Omega$ the mapping $T(\omega, \cdot)$ is a continuous function from $X$ into $R$;*

(iii) *for every couple $(\omega, x) \in \Omega \times X$ we have $|T(\omega, x)| \leqq s(\omega)$.*

PROOF: We shall essentially follow the construction in the nonprobabilistic version of this theorem as given by Alexandroff (see pp. 182–183 in [7]), the only difference being in the definition of sets $A_n(\omega)$ and $B_n(\omega)$. For the sake of definiteness we shall briefly describe the construction of the random transform $T$.

We set $V_0(\omega, x) = V(\omega, x)$ for every couple $(\omega, x) \in \Omega \times M$, and for every $n = 0, 1, 2, \cdots$ we use the following recursive formulae: For every $\omega \in \Omega$ we define

$$A_n(\omega) = \{x : V_n(\omega, x) < -(2^n/3^{n+1}) \cdot s(\omega)\}$$

and

$$B_n(\omega) = \{x : V_n(\omega, x) > (2^n/3^{n+1}) \cdot s(\omega)\}.$$

Let $\rho(x, y)$ and $\rho(x, A)$ denote the distance from the point $x$ to the point $y$ or to the set $A$. Then for every couple $(\omega, x) \in \Omega \times X$ we put (the modification in case $A_n(\omega)$ or $B_n(\omega)$ is empty is omitted) $T_n(\omega, x) = (2/3)^{n+1} \cdot s(\omega) \cdot \rho(x, A_n(\omega))/(\rho(x, A_n(\omega)) + \rho(x, B_n(\omega))) - (2^n \cdot s(\omega)/3^{n+1})$ and for every couple $(\omega, x) \in \Omega \times M$

$$V_{n+1}(\omega, x) = V_n(\omega, x) - T_n(\omega, x).$$

Finally for every couple $(\omega, x) \varepsilon \Omega \times X$ we define

$$T(\omega, x) = \lim_{n\to\infty} \sum_{k=1}^{n} T_k(\omega, x).$$

It can be easily seen that the mapping $T$ satisfies the three numbered requirements, hence we need only prove its measurability. First we shall prove that the mapping $\rho(x, A_n(\cdot))$ is for every $x \varepsilon X$ and for every $n = 0, 1, 2, \cdots$ a real-valued random variable. We have for every $c \geqq 0$ the equality

$$(1) \quad \{\omega:\rho(x, A_n(\omega)) > c\} = \bigcup_{k=1}^{\infty} \bigcap_{y\varepsilon \tilde{O}(x,k)} \{\omega: V_n(\omega, y) \geqq -(2^n \cdot s(\omega)/3^{n+1})\},$$

where $\tilde{O}(x, k)$ is a countable set dense in the set $\{y:\rho(x, y) \leqq c + (1/k)\} \cap M$. Indeed, if $\omega_0$ belongs to the set on the left hand side of (1), then there exists a positive integer $k_0$ (dependent in general on $\omega_0$) so that

$$\inf_{y\varepsilon A_n(\omega_0)} \rho(x, y) > c + (1/k_0)$$

and hence the set $A_n(\omega_0)$ and the set $\{y:\rho(x, y) \leqq c + (1/k_0)\}$ are disjoint. Therefore for this $k_0$ and for every $y \varepsilon \tilde{O}(x, k_0)$ we have

$$(2) \quad V_n(\omega_0, y) \geqq -(2^n \cdot s(\omega)/3^{n+1})$$

and this means that $\omega_0$ belongs to the set on the right hand side of (1). Conversely, let $\omega_0$ belong to the set on the right hand side of (1). Then there exists such a positive integer $k_0$, that for every $y \varepsilon \tilde{O}(x, k_0)$ the inequality (2) holds. Since the mapping $V_n(\omega_0, \cdot)$ is continuous, the inequality (2) holds for every element from the set $\{y:\rho(x, y) \leqq c + (1/k_0)\} \cap M$ and therefore the sets $A_n(\omega_0)$ and $\{y:\rho(x, y) \leqq c + (1/k_0)\}$ are disjoint. Hence

$$\inf_{y\varepsilon A_n(\omega_0)} \rho(x, y) \geqq c + (1/k_0) > c$$

and $\omega_0$ belongs also to the set on the left hand side of (1). Thus, provided $V_n$ is a random transform, we have that $\rho(x, A_n(\cdot))$ is a real-valued random variable for every $x \varepsilon X$ and quite a similar consideration holds for the mapping $\rho(x, B_n(\cdot))$. Therefore from the measurability of the mapping $V_n$ it follows that both $T_n$ and $V_{n+1}$ are also random transforms. Since $V_0$ is a random transform, the same holds for $T$. The proof is complete.

**4. Probabilistic Hahn-Banach theorem.** The next theorem forms a probabilistic version of the well-known Hahn-Banach theorem for normed linear spaces.

THEOREM 2: *Let $(\Omega, \mathfrak{S})$ be a measurable space, $X$ a separable real normed linear space, $M$ a linear manifold in the space $X$, and $V$ a random transform of the Cartesian product $\Omega \times M$ into the space $(R, \mathfrak{R})$, satisfying the following conditions:*

*for every $\omega \varepsilon \Omega$, $\alpha \varepsilon R$, $\beta \varepsilon R$, $x \varepsilon M$ and $y \varepsilon M$*

$$V(\omega, \alpha x + \beta y) = \alpha V(\omega, x) + \beta V(\omega, y);$$

*for every couple $(\omega, x) \in \Omega \times M$ we have $| V(\omega, x) | \leqq s(\omega) \cdot \| x \|$, provided the mapping $s$ of the space $\Omega$ into the space $R$ is for every $\omega \in \Omega$ defined by the formula $s(\omega) = \sup_{x \in O \cap M} | V(\omega, x) |$, where $O = \{x: \| x \| = 1\}$.*

*Then there exists a random transform $T$ of the Cartesian product $\Omega \times X$ into the space $(R, \mathfrak{R})$ so that*

(iv) *for every couple $(\omega, x) \in \Omega \times M$ we have $T(\omega, x) = V(\omega, x)$;*

(v) *for every $\omega \in \Omega$, $\alpha \in R$, $\beta \in R$, $x \in X$ and $y \in X$ there holds $T(\omega, \alpha x + \beta y) = \alpha T(\omega, x) + \beta T(w, y)$;*

(vi) *for every couple $(\omega, x) \in \Omega \times X$ we have $| T(\omega, x) | \leqq s(\omega) \cdot \| x \|$.*

PROOF: First of all we shall describe the construction of the mapping $T$.

Since $X$ is separable, there exists a countable set $\{x_1, x_2, \cdots\} \subset X - M$ dense in the set $X - M$. Let for every $n = 0, 1, 2, \cdots$ the symbol $M_n$ denote the linear manifold generated by the set $M \cup \bigcup_{k=1}^{n} \{x_k\}$ and let $X_0 = \bigcup_{n=0}^{\infty} M_n$. We set for every couple $(\omega, x) \in \Omega \times M_0$,

$$V_0(\omega, x) = V(\omega, x)$$

and for every couple $(\omega, x) \in \Omega \times (X_0 - M_0)$,

$$V_0(\omega, x) = 0.$$

Then for every $n = 1, 2, \cdots$ we define recursively for every couple $(\omega, x) \in \Omega \times (X_0 - M_n)$

$$V_n(\omega, x) = V_{n-1}(\omega, x) = 0$$

and for every $\omega \in \Omega$, $x \in M_{n-1}$ and $t \in R$

$$V_n(\omega, x + tx_n) = V_{n-1}(\omega, x) + t \cdot \sup_{x \in M_{n-1}} (V_{n-1}(\omega, x) - s(\omega) \cdot \| x - x_n \|).$$

Further we put for every couple $(\omega, x) \in \Omega \times X_0$

$$T(\omega, x) = T_0(\omega, x) = \lim_{n \to \infty} V_n(\omega, x),$$

and finally for every $y \in X - X_0$ which can be written in the form $y = \lim_{n \to \infty} y_n$, where $y_n \in X_0$ for every $n = 1, 2, \cdots$, we set

$$T(\omega, y) = \lim_{n \to \infty} T_0(\omega, y_n).$$

It is well known that for every $\omega \in \Omega$ the mapping $T(\omega, \cdot)$ is a bounded linear functional which is an extension of the bounded linear functional $V(\omega, \cdot)$ from the linear manifold $M$ to the whole space $X$ with preservation of the norm. Thus, only measurability remains to be proved. However, we can write

$$\{\omega: s(\omega) \leqq c\} = \bigcap_{x \in \tilde{O}} \{\omega: | V_n(\omega, x) | \leqq c\},$$

where $\tilde{O}$ is a countable set dense in the set $O$. Since $V$ is a random transform, the mapping $V_0$ is a random transform of the Cartesian product $\Omega \times X_0$ into the

space $(R, \mathfrak{R})$. Further we have for every $c \varepsilon R$

$$\{\omega: \sup_{x \varepsilon M_{n-1}} (V_{n-1}(\omega, x) - s(\omega) \cdot \| x - x_n \|) \leqq c\}$$
$$= \bigcap_{x \varepsilon \tilde{M}_{n-1}} \{\omega: V_{n-1}(\omega, x) - s(\omega) \cdot \| x - x_n \| \leqq c\},$$

where $\tilde{M}_{n-1} \subset M_{n-1}$ is a countable set dense in the set $M_{n-1}$. Thus, $T_0$ is a random transform of the Cartesian product $\Omega \times X_0$ into the space $(R, \mathfrak{R})$ and therefore the mapping $T$ is a random transform of the Cartesian product $\Omega \times X$ into the space $(R, \mathfrak{R})$ and this proves Theorem 2.

Theorem 2 states that for separable Banach spaces a probabilistic version of the Hahn-Banach theorem is valid. Theorem 3 below shows that for (not necessarily separable) Hilbert spaces an equivalent statement is also true.

**5. Conclusion.** It would be of interest to know the extent to which Theorems 1 and 2 remain valid if the separability assumption is dropped. In this case the methods of proof used above obviously fail. Unfortunately, the author has not succeeded either in proving the non-separable versions or in constructing appropriate counterexamples. To get other positive results it seems necessary to lay further assumptions on the space $X$. Thus, a statement equivalent to Theorem 2 is true for not necessarily separable Hilbert spaces, owing to the possibility of defining an orthogonal complement to a given subspace.

THEOREM 3: *Theorem 2 remains valid provided $X$ is a Hilbert space and $M$ a Hilbert subspace of the space $X$.*

PROOF. Since every element $x \varepsilon X$ can be uniquely written in the form $x = z_1 + z_2$, where $z_1 \varepsilon M$ and $z_2 \perp M$, we can set for every $\omega \varepsilon \Omega$

$$T(\omega, x) = V(\omega, z_1)$$

and Theorem 3 follows immediately from this construction.

Finally, let us briefly sketch an application of our results.

The well-known Banach-Mazur Theorem asserts that every separable metric (Banach) space $M$ can be imbedded in an isometric (isometric and isomorphic) way into the space $C$ of all continuous functions defined on the closed interval $(0, 1)$. This theorem enables us sometimes to treat generalized random variables with values in the space $(M, \mathfrak{M})$ as generalized random variables with values in the space $(C, \mathfrak{C})$. This is the case in Theorem 16 in [1], where the measurability of the set $\{\omega: \bigcup_{n=1}^{\infty} \{V_n(\omega)\}$ is strongly compact$\}$ must be proved. Another example is Criterion 6 in [1] (for wording see Introduction of this paper). In both these cases the above mentioned treatment provides a simple and elegant proof of the statement in question.

Using Theorems 1 and 2 we are able to enlarge the number of problems in which not only generalized random variables with values in the space $(M, \mathfrak{M})$ are considered, but also random transforms of the Cartesian product $\Omega \times M$ into the space $(R, \mathfrak{R})$. In the present paper we shall mention only one problem of this kind, namely the Representation Theorem for random Schwartz distribu-

tions which, roughly speaking, reads: Every random Schwartz distribution can be represented on every compact interval with arbitrarily great probability as a derivative of a strictly continuous stochastic process. This theorem was proved by Ullrich [8] who applied our Theorem 2 with $X$ replaced by $C$ and $M$ by $K_r$, where $K_r$ stands for the space of $r$th derivatives of all continuous functions $f$ defined on a closed interval $[a, b]$ that have derivatives of all orders, the functions $f$ themselves and their derivatives taking the value 0 at both ends of the interval $[a, b]$.

## REFERENCES

[1] O. Hanš, "Generalized random variables," *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague 1957, pp. 61–103.

[2] E. Mourier, "Éléments aléatoires dans un espace de Banach," *Ann. Inst. H. Poincaré*, Vol. 13 (1953), pp. 161–244.

[3] J. Nedoma, "Note on generalized random variables," *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*; Prague 1957, pp. 139–141.

[4] A. T. Bharucha-Reid, "On random elements in Orlicz spaces," *Bull. Acad. Poll.*, Cl III, Vol. 4 (1956), pp. 665–667.

[5] A. N. Kolmogoroff und Ju. W. Prochorow, "Zufällige Funktionen und Grenzverteilungssätze," *Bericht über die Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik*; Berlin 1956, pp. 113–126.

[6] L. E. Dubins, "Generalized random variables," *Trans. Amer. Math. Soc.*, Vol. 84 (1957), pp. 273–309.

[7] P. S. Alexandroff, "*Einführung in die Mengenlehre und die Theorie der reellen Funktionen*," Berlin, 1956.

[8] M. Ullrich, "Representation Theorem for random Schwartz distributions," *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague, 1960.

# A CONVOLUTIVE CLASS OF MONOTONE LIKELIHOOD RATIO FAMILIES

By S. G. Ghurye and David L. Wallace

*University of Chicago*

**1. Introduction.** This note stems from the following problem posed us by J. Loevinger.[1] Let $X_1, \cdots X_n$ and $Y$ be real-valued random variables such that, conditionally on $Y$, the $\{X_i\}$ are mutually independent with

$$p_i(Y) = \Pr\{X_i = 1 \mid Y\} = 1 - \Pr\{X_i = 0 \mid Y\}$$

and $p_i(y)$ is nondecreasing in $y$. Let $S = X_1 + \cdots + X_n$. Is $E\{Y \mid S = r\}$ a nondecreasing function of $r$? The answer, yes, will follow from showing that $\Pr\{S = r + 1 \mid Y\}/\Pr\{S = r \mid Y\}$ is a nondecreasing function of $Y$ for each $r$. Here we have a simple case of the convolution of families of distributions with monotone likelihood ratios (hereafter MLR) being an MLR family. It is easy to see that the convolution of two MLR families is not necessarily MLR. In Section 2, a sufficient condition on MLR families is given that their convolution be MLR. In Section 3, some special results are given for multidimensional distributions. The problem leading to this work is discussed in Section 4.

The MLR property is identical with the Pólya type 2 property (cf. [2]). The definitions used here extend to Pólya type $m$ but the extended results, except for Lemma 4, are not generally true for $m > 2$.

**2. Convolutions of MLR families.** Let $G$ be an ordered additive group, let $\Theta$ be an ordered set, and let $\mu$ be an invariant, $\sigma$-finite measure on $G$. Throughout this section, a family $f$ will mean a real-valued, nonnegative function on $G \times \Theta$, such that $f(x, \theta)$ is measurable in $x$ for each $\theta$ and

$$0 < \textstyle\int_G f(x, \theta)\, d\mu(x) < \infty.\text{[2]}$$

Ordinarily, $f$ is a family of probability densities relative to $\mu$ for a random variable with range contained in $G$ and with parameter space $\Theta$. The convolution family $f * g$ of two families $f$ and $g$ is defined by

$$f * g(x, \theta) = \int_G f(x - u, \theta) g(u, \theta)\, d\mu(u).$$

The spaces $G$ and $\Theta$ must be ordered for the definitions which follow and $G$ must be at least a semigroup for convolutions to be defined in the same space.

Received January 14, 1959; revised April 15, 1959.

[1] The consultation with Dr. Loevinger (Jewish Hospital of St. Louis) and the research connected with this paper were performed under a grant from the Rockefeller Foundation.

[2] For some purposes, such as in Section 3, it would be convenient to permit the integral to be zero for some $\theta$. Lemma 3 and the theorems of this section clearly hold under this extension.

The group requirement is only a slight restriction. $G$ will ordinarily be the real line or the integers, and is taken as an ordered group primarily because it permits a simple unified treatment at no extra cost.

*Definition.* A nonnegative function $h$ defined on the product of two ordered sets $X$, $Y$, is Pólya type 2 if, for all $x, x' \varepsilon X$ and $y, y' \varepsilon Y$ such that $x \leqq x'$, $y \leqq y'$,

$$h(x, y)h(x', y') - h(x, y')h(x', y) \geqq 0.$$

*Definition A.* A family $f$ has property $A$ if, as a function on $G \times \Theta$, it is Pólya type 2.

*Definition B.* A family $f$ has property $B$ if, for each $\theta \varepsilon \Theta$, the function $h_\theta$ on $G \times G$ defined by

$$h_\theta(x, \xi) = f(x - \xi, \theta)$$

is Pólya type 2.

*Definition C.* A family $f$ has property $C$ if, for all $x, x', y, y' \varepsilon G$ such that $y \leqq x \leqq y'$ and $x + x' = y + y'$ and all $\theta, \theta' \varepsilon \Theta$ such that $\theta \leqq \theta'$,

$$f(x, \theta)f(x', \theta') \geqq f(y', \theta)f(y, \theta').$$

Property $A$ is the monotone likelihood ratio or Pólya type 2 property for the family $f$. Property $B$ is the monotone likelihood ratio property for the location parameter family generated by $f(\cdot, \theta)$ for each fixed $\theta$. Provided all quantities used as divisors are positive, the definitions of properties $A$ and $B$ can be expressed in the more intuitive form:

$A: f(x, \theta')/f(x, \theta)$ nondecreasing in $x$ for all $\theta < \theta'$, or

$A: f(x + h, \theta)/f(x, \theta)$ nondecreasing in $\theta$ for all $x$ and all $h > 0$, and

$B: f(x + h, \theta)/f(x, \theta)$ nonincreasing in $x$ for all $\theta$ and all $h > 0$.

Note that on taking $x = y$ and $x' = y'$ in $C$, one obtains $A$; that on taking $\theta = \theta'$ in $C$, one obtains $B$. We shall now show that property $C$ is, in fact, equivalent to $A$ and $B$ together, and that it is invariant under convolution.

It may be helpful to note that all results and methods of this paper are unaffected if any $f(x, \theta)$ is multiplied by any positive function of $\theta$. Multiplication by a positive function of $x$ does not destroy MLR, but does affect the convolution and its MLR properties.

LEMMA 1: *If $f$ has property $B$, then the set $I_f(\theta) = \{x: f(x, \theta) > 0\}$ is, for every $\theta \varepsilon \Theta$, an interval of $G$; i.e., $y \varepsilon I$, $y' \varepsilon I$ imply $x \varepsilon I$ for all $x \varepsilon G$ such that $y \leqq x \leqq y'$.*

PROOF: Suppose $f(y, \theta)f(y', \theta) > 0$ and $y < y'$. To any $x \varepsilon G$ such that $y \leqq x \leqq y'$, there corresponds an $x' \varepsilon G$ such that $x + x' = y + y'$ and by property $B$, $f(x, \theta)f(x', \theta) \geqq f(y, \theta)f(y', \theta) > 0$.

Thus, for each $\theta$, there is a decomposition of $G$ into three intervals $M(\theta)$, $I(\theta)$, $M'(\theta)$ such that $x \varepsilon M$, $y \varepsilon I$, $z \varepsilon M'$ imply $x < y < z$ and $f(x, \theta) = f(z, \theta) = 0$, $f(y, \theta) > 0$. For all $\theta$, $I_f(\theta)$ is nonempty, though it may contain only one point. $M(\theta)$ and $M'(\theta)$ may be empty.

LEMMA 2: *If $f$ has properties A and B, then for any $\theta$, $\theta' \varepsilon \Theta$ such that $\theta < \theta'$, $M(\theta) \subset M(\theta')$ and $M'(\theta) \supset M'(\theta')$.*

PROOF: For any $\theta$, choose $x' \varepsilon I(\theta)$. Then for any $x \varepsilon M(\theta)$, $x < x'$ and using property $A$, $f(x, \theta') = 0$. Since this holds for all $y \leqq x$, then $x \varepsilon M(\theta')$. A similar proof holds for $M'(\theta)$.

LEMMA 3: *$f$ has property C if and only if it has properties A and B.*

PROOF. That $C$ implies $A$ and $B$ is immediate. Property $C$ is nontrivial only when $y' \ \varepsilon \ I(\theta)$ and $y \ \varepsilon \ I(\theta')$. Since $y < y'$, we know from Lemma 2 that $f(y, \theta) > 0$. Using successively $A$ and $B$,

$$f(y, \theta)f(x, \theta)f(x', \theta') \geqq f(y, \theta')f(x, \theta)f(x', \theta) \geqq f(y, \theta')f(y, \theta)f(y', \theta)$$

and $C$ follows by division.

LEMMA 4. (Schoenberg [5]). *If $f$ and $g$ have property B, then $f * g$ has property B.*

Schoenberg's proof for the real line extends immediately to the group $G$. He proves this result in its Pólya type $m$ form.

THEOREM 1: *If $f$ and $g$ have property C, then $f * g$ has property C.*

PROOF: Using Lemmas 3 and 4, it remains only to show that $f * g$ has property $A$, i.e., for $x < x' = x + h$, $\theta < \theta'$,

$$\Delta_2 \equiv [f * g(x, \theta)][f * g(x', \theta')] - [f * g(x', \theta)][f * g(x, \theta')] \geqq 0.$$

Throughout the proof, write $f(x) = f(x, \theta)$ and $f'(x) = f(x, \theta')$.

$$\Delta_2 = \int [f(u)g(x - u)f'(v)g'(x' - v) - f'(u)g'(x - u)f(v)g(x' - v)] \cdot d[\mu(u) \times \mu(v)] = I_1 + I_2 + I_3$$

in which $I_1, I_2, I_3$ are respectively the integrals over the sets, $u > v, u = v, u < v$.

Interchange $u$ and $v$ in $I_3$ and incorporate with $I_1$.

$$I_1 + I_3 = \int_{u>v} \{f(u)f'(v)[g(x - u)g'(x' - v) - g(x' - u)g'(x - v)] + f'(u)f(v)[g(x - v)g'(x' - u) - g'(x - u)g(x' - v)]\} \cdot d[\mu(u) \times \mu(v)].$$

For $u > v$, the quantity in the second brackets is nonnegative by $C$ and its coefficient $f'(u)f(v) \geqq f(u)f'(v)$. Then,

$$I_1 + I_3 \geqq \int_{u>v} f(u)f'(v)[g(x - u)g'(x' - v) - g'(x - u)g(x' - v) + g(x - v)g'(x' - u) - g(x' - u)g'(x - v)]\, d[\mu(u) \times \mu(v)]$$

and

$$\Delta_2 \geqq \int_{u>v} f(u)f'(v)[g(x - u)g'(x' - v) - g'(x - u)g(x' - v)]\, d[\mu(u) \times \mu(v)] + \int_{u>v+2h} f(u)f'(v)[g(x - v)g'(x' - u) - g(x' - u)g'(x - v)]\, d[\mu(u) \times \mu(v)]$$

$$+ \int_{0 \leq u-v \leq 2h} f(u)f'(v)[g(x-v)g'(x'-u) - g(x'-u)g'(x-v)] \cdot d[\mu(u) \times \mu(v)]$$

$= J_1 + J_2 + J_3$ respectively.

In $J_2$ make the transformation $u = u' + h$, $v = v' - h$, suppress the primes on $u'$, $v'$, and recall that $h = x' - x > 0$. Then

$$J_1 + J_2 = \int_{u>v} [f(u)f'(v) - f(u+h)f'(v-h)] \cdot [g(x-u)g'(x'-v) - g'(x-u)g(x'-v)]\, d[\mu(u) \times \mu(v)] \geq 0.$$

Break $J_3$ into three integrals respectively over the sets $0 \leq u - v < h$, $h < u - v \leq 2h$, $u - v = h$. The third integral vanishes, and on making the transformation $u = v' + h$, $v = u' - h$ in the second and suppressing the primes on $u'$, $v'$, we get

$$J_3 = \int_{0 \leq u-v<h} [f(u)f'(v) - f(v+h)f'(u-h)] \cdot [g(x-v)g'(x'-u) - g(x'-u)g'(x-v)]\, d[\mu(u) \times \mu(v)] \geq 0.$$

Hence $\Delta_2 \geq 0$, and the theorem is proved.

*Remark.* This result would appear to be subsumed under Theorem 3 of Lehmann [3], taking $g_\theta(x, \xi) = f(x - \xi, \theta)$ and $d\lambda_\theta(\xi) = g(\xi, \theta)\, d\mu(\xi)$. However, in lines 6 to 8 of page 410 of his proof, an additional assumption is needed, which is not met in our case.[3]

COROLLARY 1: *If $\Theta = G$, $f(x, \theta) = f(x - \theta)$, $g(x, \theta) = g(x - \theta)$ and $f$ and $g$ have property A, then $f * g$ has property A.* (This result for location parameter families is known and is just the Schoenberg result of Lemma 4.)

COROLLARY 2: *If $G$ is the set of integers, if for each $i = 1, \cdots, n$,*

$$f_i(x, \theta) = \begin{cases} p_i(\theta) & x = 1 \\ 1 - p_i(\theta) & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

*and $p_i(\theta)$ is nondecreasing in $\theta$, then each $f_i$ and the convolution $f_1 * \cdots * f_n$ have property A.*

That $B$ is not a necessary property for the convolution of two MLR families to be MLR is shown by the construction below based on the following theorem, whose proof is a simple computation.

THEOREM 2: *If $f$ has property A and if, for each $\theta$, the range of $x$ for which $f(x, \theta) > 0$ is contained in $(0, 1, 2)$, then $f * f$ has property A.*

This result does not extend in general to nonidentical convolutions, to threefold identical convolutions, or to fourpoint ranges.

A family $f$ which satisfies Theorem 2 but does not have property $B$ is easily

[3] We wish to thank Professor S. Karlin for calling this fact to our attention.

constructed by taking $\Theta$ as the real line, $a(\theta)$, $b(\theta)$ as increasing functions on $\Theta$ such that $0 < a(\theta) < b(\theta) < \infty$, and letting $f(\cdot, \theta)$ be the distribution with probabilities at 0, 1 and 2 respectively given by

$$c(\theta), \qquad a(\theta)c(\theta), \qquad a(\theta)b(\theta)c(\theta)$$

with $c(\theta) = [1 + a(\theta) + a(\theta)b(\theta)]^{-1}$.

**3. Some results for multivariate distributions.** A family of generalized densities $f(x, \theta)$, where $x$ is a vector is said to be MLR (or Pólya type 2) if it is MLR along each increasing curve, i.e., if for every vector function $x(t)$ of the real-parameter $t$ for which the components are nondecreasing functions of $t$, $g(t, \theta) = f\{x(t), \theta\}$ is MLR in $t$ and $\theta$. (Cf. Lehmann [3], Pratt [4].) The definition can also be stated in the form:

$f(x_1, \cdots, x_K, \theta)$ is MLR if, for all $x_i \leqq x_i'$, $i = 1, \cdots, K$, $\theta \leqq \theta'$,

$$f(x_1, \cdots, x_K, \theta)f(x_1', \cdots, x_K', \theta') \geqq f(x_1', \cdots, x_K', \theta)f(x_1, \cdots, x_K, \theta').$$

We consider only the simplest problem of extending Corollary 2 to families of distributions on the vertices of the cube or the simplex in $K$ dimensions. In two dimensions already, two MLR families on the points (0, 0), (0, 1), (1, 0) need not have an MLR convolution (Counterexample 1). Restricting consideration to $n$-fold convolutions of a single family, the $n$-fold convolution of an MLR family on the vertices of the square is MLR (Theorem 3), but even the two-fold convolution of an MLR family on the vertices of the three-dimensional cube need not be MLR (Counterexample 2). However, the $n$-fold convolution of an MLR family on the vertices of the $K$-dimensional simplex is MLR for all $n$ and $K$ (Theorem 2).

*Counterexample* 1. The convolution of two MLR families $f_1$ and $f_2$ on the points (0, 0), (1, 0), (0, 1) need not be MLR: Let $a(\theta)$ be a positive, increasing function of $\theta$ and let $f_1$ place nonzero probabilities only on the three points (0, 0), (1, 0), (0, 1) proportional, respectively, to 1, 2, $a(\theta)$. Let $f_2(x, \theta) = \frac{1}{3}$ at each point. Both are MLR. Then at the two points (0, 1) and (1, 1), $f_1 * f_2$ has probabilities proportional, respectively, to $[1 + a(\theta)]$ and $[2 + a(\theta)]$, and hence $f_1 * f_2$ is not MLR.

*Counterexample* 2. The two-fold convolution of an MLR family on the vertices of the three-dimensional cube need not be MLR: Let $f$ place nonzero probabilities only on the five points (0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0) proportional, respectively, to 1, $a(\theta)$, $a(\theta)$, 2, $a(\theta)$, with $a(\theta)$ as above. Then $f$ is MLR but $f * f$ is not, since at the points (1, 1, 0) and (1, 1, 1) the probabilities are proportional, respectively, to $[1 + a(\theta)]$ and 2.

THEOREM 3: *If $f$ is an MLR family on the four points* (0, 0), (1, 0), (0, 1), (1, 1), *then, for every $n$, the $n$-fold convolution of $f$ with itself is MLR.*

PROOF. Let $f(x, \theta) = p_{ij}(\theta)$, for $x = (i, j)$, and let $q_{ij}(\theta)$ be the value of the $n$-fold convolution of $f$ at $(i, j)$, which is given by

$$(1) \qquad \sum_{i,j} q_{ij}(\theta)t^i u^j = [\sum_{i,j} p_{ij}(\theta)t^i u^j]^n.$$

We are given that for all $i \leqq i', j \leqq j', \theta \leqq \theta'$,

$$p_{ij}(\theta)p_{i'j'}(\theta') \geqq p_{ij}(\theta')p_{i'j'}(\theta), \tag{2}$$

and must show that for all $r \leqq r'$, $s \leqq s'$, $\theta \leqq \theta'$,

$$q_{rs}(\theta)q_{r's'}(\theta') \geqq q_{rs}(\theta')q_{r's'}(\theta). \tag{3}$$

From (1), it follows that, for given $s$, the sequence $\{q_{rs}(\theta), r = 0, 1, \cdots\}$ has the generating function

$$P_s(t) = \binom{n}{s}\{p_{00}(\theta) + p_{10}(\theta)t\}^{n-s}\{p_{01}(\theta) + p_{11}(\theta)t\}^s$$

and, for given $r$, the sequence $\{q_{rs}(\theta), s = 0, 1, \cdots\}$ has the generating function

$$Q_r(u) = \binom{n}{r}\{p_{00}(\theta) + p_{01}(\theta)u\}^{n-r}\{p_{10}(\theta) + p_{11}(\theta)u\}^r.$$

Both represent convolutions of two-point, one-dimensional families, MLR by (2), and hence, by Corollary 2 to Theorem 1,

$$q_{rs}(\theta)q_{r's}(\theta') \geqq q_{r's}(\theta)q_{rs}(\theta')$$

and

$$q_{rs}(\theta)q_{rs'}(\theta') \geqq q_{rs'}(\theta)q_{rs}(\theta').$$

The desired conclusion (3) follows easily if at least one of $q_{r's}(\theta)$, $q_{r's}(\theta')$, $q_{rs'}(\theta)$, $q_{rs'}(\theta')$ is positive. (3) is trivially true unless $q_{rs}(\theta')q_{r's'}(\theta) > 0$ and is one of the above special cases unless $r' > r$ and $s' > s$. If $r' < s'$, then $q_{r's'}(\theta) > 0$ implies either $p_{11}(\theta)p_{01}(\theta) > 0$ or $p_{10}(\theta)p_{01}(\theta) > 0$. In either case, $q_{rs'}(\theta) > 0$ and (3) follows. A similar argument holds if $r' > s'$ and also if $r' = s'$ except when $p_{11}(\theta) > 0$, $p_{10}(\theta) = p_{01}(\theta) = 0$. But then, by the MLR property, also $p_{10}(\theta') = p_{01}(\theta') = 0$ and at $\theta, \theta'$ the distributions are one-dimensional along the diagonal. Hence Corollary 2 (for the group of diagonal integers) applies directly.

THEOREM 4: *If $f$ is an MLR family on the $K + 1$ vertices of the unit simplex in $K$ dimensions, then, for every $n$, the $n$-fold convolution of $f$ with itself is MLR.*

PROOF: Let $\{p_j(\theta), j = 0, 1, \cdots K\}$ be the values of the family $f$ at the origin and unit points of the $K$ axes respectively. Let $q_r(\theta)$ denote the value of the $n$-fold convolution at the point $r = (r_1, \cdots, r_K)$. We must show that for $\theta \leqq \theta'$, and $r, r'$ such that $r_i \leqq r_i'$, $i = 1, \cdots, K$,

$$q_r(\theta)q_{r'}(\theta') \geqq q_r(\theta')q_{r'}(\theta). \tag{4}$$

A generating function argument similar to that used above easily proves the result when $r$ and $r'$ differ in only one coordinate. But if $q_{r'}(\theta) > 0$, then $p_j(\theta) > 0$ for all coordinates such that $r_j' > 0$, and hence $q_s(\theta) > 0$ for all $s$ such that $s_j \leqq r_j'$, $j = 1, \cdots, K$. Division by $q_s(\theta)$ is permissible and (4) follows easily by repeated application of the result for changes in a single coordinate.

**4. An application.** The problem mentioned in the introduction arose in the following way. Let $X_1, \cdots, X_n$ denote the scores made by an individual on

$n$ test items with the value 1 for correct, 0 for incorrect. Let $S = X_1 + \cdots + X_n$. Let $Y$ be a real random variable representing the individual's (unobservable) position on a single scale (latent continuum) assumed to determine his performance on the test according to the model: For each $i$,

$$f_i(1, Y) = p_i(Y) = \Pr\{X_i = 1 \mid Y\} = 1 - \Pr\{X_i = 0 \mid Y\}$$

and conditionally on $Y$, the $\{X_i\}$ are mutually independent. Let $\phi$ be the probability density function for $Y$, representing, perhaps, the distribution of the ability $Y$ over some population. If it is assumed only that each $f_i$ is a nondecreasing function, what can be said about the individual value of $Y$ conditionally on the sum $S$ of the scores of the $n$ items? The answer is that the conditional distribution function of $Y$ given $S = a$ lies to the left of that for $S = b > a$. Hence, the conditional mean (or median or quantile) of $Y$ given $S$ is a nondecreasing function of $S$.

(The result would not be true without restrictions on the functions $p_i$ if, for example, the difference between a correct and an incorrect score differed from item to item.)

The result is an application of Corollary 2 to Theorem 1 and of the following lemma.

LEMMA. *If $X$, $Y$ are real random variables, if $Y$ has density $\phi(y)$ relative to the measure $\nu$, if $X$ given $Y = y$ has the conditional density $f(x, y)$ relative to the measure $\mu$, and if the family $f$ is MLR, then*

$$\Pr\{Y \leqq a \mid X = x\} \geqq \Pr\{Y \leqq a \mid X = x'\}$$

*for all $a$ and all $x \leqq x'$.*

The lemma can be proved by relatively simple calculations and is equivalent to a result of Good [1].

## REFERENCES

[1] I. J. GOOD, "A note on positive determinants," *J. London Math Soc.*, Vol. 22 (1947), pp. 92–95.

[2] SAMUEL KARLIN, "Decision theory for Pólya type distributions. Case of two actions, I," *Proceedings of the Third Berkeley Symposium on Probability and Statistics*, Vol. 1, University of California Press, 1955, pp. 115–128.

[3] E. L. LEHMANN, "Ordered families of distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 399–419.

[4] J. W. PRATT, "Some results in the decision theory of one-parameter multivariate Pólya type distributions," *Stanford Technical Report No. 37*, 21 October 1955.

[5] I. J. SCHOENBERG, "On Pólya frequency functions, I. The totally positive functions and their Laplace transforms," *J. d'Analyse Math.*, Vol. 1 (1951), pp. 331–374.

# ON THE LAWS OF CAUCHY AND GAUSS

By R. G. Laha[1]

*The Catholic University of America*

**1. Introduction and summary.** Let $x$ and $y$ be two independent normal variates each distributed with zero mean and a common variance; it is then well-known that the quotient $x/y$ follows the Cauchy law distributed symmetrically about the origin. Now the question that naturally arises is whether we can obtain a characterization of the normal distribution by this property of the quotient. This converse problem can be more precisely formulated as follows:

Let $x$ and $y$ be two independently and identically distributed random variables having a common distribution function $F(x)$. Let the quotient $w = x/y$ follow the Cauchy law distributed symmetrically about the origin $w = 0$. Then the question is whether $F(x)$ is normal.

But this converse is not true in general. The author [1] has recently constructed a very simple example of a non-normal distribution where the quotient $x/y$ follows the Cauchy law. Steck [7] has also given some examples of non-normal distributions with this property of the quotient.[2]

In the present paper we shall first derive some interesting general properties possessed by the class of distribution laws $F(x)$ [Section 2]. In Section 3 we deduce a characterization of the normal distribution under some conditions on the distribution function $F(x)$. Finally in Section 4 we construct an example of a non-normal distribution function $F(x)$ having finite moments of all orders where the quotient $x/y$ follows the Cauchy law. The method of proof is essentially based on the applications of Fourier transforms of distribution functions. For the proof of Theorem 3.1 we require somewhat deeper results in the theory of analytic functions.

**2. Some general properties of $F(x)$.** We shall here discuss some general properties of the class of distribution laws $F(x)$. We first prove a lemma which is instrumental in the proofs of the subsequent results.

Lemma 2.1. *Let $x$ and $y$ be two independently and identically distributed proper random variables having a common distribution function $F(x)$ which is continuous at the origin $x = 0$. Let the quotient $w = x/y$ have a distribution function $G(w)$ symmetric about the origin. Then $F(x)$ is also symmetric about the origin.*

Proof. As usual we assume that each of the distribution functions $F(x)$ and

---

Received March 14, 1958; revised March 3, 1959.

[1] This work was supported by the National Science Foundation through grant NSF-G-4220.

[2] *Note added in proof*: While this paper was in press, the author learned that some examples of non-normal distributions have also been constructed by J. G. Mauldon (*Quart. J. Math., Oxford* (*2*), Vol. 7 (1956), pp. 155–160).

$G(w)$ is everywhere continuous to the right. Then we have the following notations:

$$F(a) = \text{Prob}\,(x \leqq a) \qquad G(a) = \text{Prob}\,(w \leqq a)$$

$$F(a-0) = \text{Prob}\,(x < a) \qquad G(a-0) = \text{Prob}\,(w < a).$$

We also note that the origin $x = 0$ must be a continuity point of $F(x)$, as otherwise the quotient $w$ assumes the indeterminate value 0/0 with a positive probability. Now for $w > 0$ we have

$$\begin{aligned} G(w) - G(0) &= \text{Prob}\left[0 < \frac{x}{y} \leqq w\right] \\ &= \text{Prob}\,[0 < x \leqq wy; y > 0] \\ &\qquad + \text{Prob}\,[wy \leqq x < 0; y < 0] \\ &= \int_0^\infty [F(wy) - F(0)]\,dF(y) \\ &\qquad + \int_{-\infty}^0 [F(0) - F(wy - 0)]\,dF(y) \\ &= \int_0^\infty [F(wy) - F(0)]\,dF(y) \\ &\qquad + \int_0^\infty [F(-wy - 0) - F(0)]\,dF(-y - 0). \end{aligned} \tag{2.1}$$

Similarly we can show that for any $w > 0$

$$\begin{aligned} G(0) - G(-w - 0) &= \text{Prob}\left[-w \leqq \frac{x}{y} < 0\right] \\ &= \int_0^\infty [F(0) - F(-wy - 0)]\,dF(y) \\ &\qquad + \int_0^\infty [F(0) - F(wy)]\,dF(-y - 0). \end{aligned} \tag{2.2}$$

Since $G(w)$ is symmetric about the origin $w = 0$, we have the relation

$$G(w) - G(0) = G(0) - G(-w - 0) \tag{2.3}$$

holding for all $w$.

Then using (2.1) and (2.2) together, we get from (2.3) the relation

$$\begin{aligned} &\int_0^\infty [F(wy) + F(-wy - 0) - 2F(0)]\,dF(y) \\ &\qquad + \int_0^\infty [F(wy) + F(-wy - 0) - 2F(0)]\,dF(-y - 0) = 0 \end{aligned} \tag{2.4}$$

holding for all $w > 0$.

Substituting

$$H(wy) = F(wy) + F(-wy - 0) - 2F(0)$$

in (2.4), we obtain

$$\int_0^\infty H(wy)\, dH(y) = 0 \tag{2.5}$$

holding for all $w > 0$. Here $H(y)$ is a function of bounded variation. We now use the transformation $w = e^u$ and $y = e^v(-\infty \leqq u \leqq \infty; -\infty \leqq v \leqq \infty)$ and denote

$$H(y) = H(e^v) = H_1(v) \quad \text{and} \quad H(wy) = H(e^{u+v}) = H_1(u + v).$$

Here we note that $H_1(v)$ is also a function of bounded variation. Thus (2.5) reduces to

$$\int_{-\infty}^\infty H_1(u + v)\, dH_1(v) = 0 \tag{2.6}$$

holding for all $u(-\infty \leqq u \leqq +\infty)$. From (2.6) we see easily that the relation

$$\int_{-\infty}^\infty e^{iut}\, d\left[\int_{-\infty}^\infty H_1(u + v)\, dH_1(v)\right] = 0 \tag{2.7}$$

holds identically for all real $t$. Let

$$\psi(t) = \int_{-\infty}^\infty e^{ivt}\, dH_1(v) \tag{2.8}$$

denote the Fourier transform of $H_1(v)$ which is a function of bounded variation. Then using the theorem of Fourier transforms of convolutions of functions of bounded variation we get from (2.7)

$$\psi(t)\psi(-t) = |\psi(t)|^2 = 0;$$

that is,

$$|\psi(t)| = 0 \tag{2.9}$$

holding identically for all real $t$, where $\psi(t)$ is defined in (2.8). Finally from the uniqueness property of Fourier transforms of functions of bounded variation, it follows immediately from (2.9) that $H_1(v)$ is a constant almost everywhere. Hence

$$H(y) = F(y) + F(-y - 0) - 2F(0) = c, \quad \text{a.e.} \tag{2.10}$$

Next substituting $y = 0$ in (2.10) and noting that the origin $y = 0$ is a continuity point of $F(y)$, we get $c = 0$ and thus (2.10) reduces to

$$F(y) + F(-y - 0) = 2F(0). \tag{2.11}$$

Finally we note $F(-\infty) = 0$ and $F(+\infty) = 1$ and obtain from (2.11) that $F(0) = \frac{1}{2}$. Thus we have

$$F(y) = 1 - F(-y - 0), \tag{2.12}$$

which completes the proof.

**Lemma** 2.2. *Let $x$ and $y$ be two independently and identically distributed random variables having a common distribution function $F(x)$. Let the quotient $w = x/y$ follow the Cauchy law distributed symmetrically about the origin $w = 0$. Then $F(x)$ is absolutely continuous and has a continuous probability density function $f(x) = F'(x) > 0$.*

**Proof.** As a direct consequence of Lemma 2.1 it follows that $F(x)$ is also symmetric about the origin $x = 0$. Let $F_0(x)$ denote the distribution function of $|x|$. Then we can verify easily that

$$F_0(x) = \begin{cases} 0 & x < 0 \\ F(x) - F(-x - 0) = 2F(x) - 1 \quad \text{for} & x \geqq 0. \end{cases} \tag{2.13}$$

Thus we note that in this case the distribution functions of $x$ and $w$ are uniquely determined by the distribution functions of $|x|$ and $|w|$ respectively. We can easily verify after elementary integration that the characteristic function of the distribution of $\ln |w|$ is given by

$$E(e^{it\ln|w|}) = \frac{1}{\cosh\left(\frac{\pi}{2}t\right)}.$$

Then noting that $\ln |w| = \ln |x| - \ln |y|$ we get finally the relation

$$\varphi(t)\varphi(-t) = \frac{1}{\cosh\left(\frac{\pi}{2}t\right)} \tag{2.14}$$

holding for all real $t$, where $\varphi(t)$ denotes the characteristic function of the distribution of $\ln |x|$. The relation (2.14) has also been derived independently by Steck [7]. From (2.14) we get at once

$$|\varphi(t)| = \frac{1}{\left[\cosh\left(\frac{\pi}{2}t\right)\right]^{1/2}} \tag{2.15}$$

and then verify easily that $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$; that is, the characteristic function $\varphi(t)$ is absolutely integrable. Then using the well-known theorem ([2], p. 188), we deduce easily that the distribution function of $\ln |x|$ is absolutely continuous and has a continuous probability density function. Thus it follows as an immediate consequence that $|x|$ has an absolutely continuous distribution function. Finally from the relation (2.13) we see easily that $F(x)$ is also absolutely continuous and has a continuous probability density function.

We are now in a position to prove the following theorem.

THEOREM 2.1. *Let $x$ and $y$ be two independently and identically distributed random variables having a common distribution function $F(x)$. Let the quotient $w = x/y$ follow the Cauchy law distributed symmetrically about the origin $w = 0$. Then $F(x)$ has the following general properties:*

(1) *it is symmetric about the origin $x = 0$;*

(2) *it is absolutely continuous and has a continuous probability density function $f(x) = F'(x) > 0$;*

(3) *the random variable $x$ has an unbounded range;*

(4) *the probability density function $f(x)$ satisfies the integral equation*

$$\int_0^\infty f(x)f(ux)x\,dx = \frac{c_0}{1+u^2} \tag{2.16}$$

*holding for all $u$, where $c_0$ is a constant.*

PROOF. The properties (1) and (2) follow as direct consequences of Lemmas 2.1 and 2.2. For the proof of property (3) we proceed as follows:

Let us suppose that the random variable $x$ has a bounded range, that is, $F(x)$ is contained in a finite interval $(-a, +a)$ of the $x$-axis. We introduce the polar transformation $x = r \cos \theta$ and $y = r \sin \theta$ and deduce easily that the joint probability density function of $r$ and $\theta$ has the form

$$r\,f(r \cos \theta)f(r \sin \theta). \tag{2.17}$$

We now integrate (2.17) with respect to $r$ and obtain the probability density function of $\theta$ as:

$$\begin{aligned} f_1(\theta) &= \int_0^{a/\cos\theta} f(r \cos \theta)f(r \sin \theta)r\,dr \qquad \text{for} \quad 0 \leqq \theta \leqq \pi/4 \\ f_2(\theta) &= \int_0^{a/\sin\theta} f(r \cos \theta)f(r \sin \theta)r\,dr \qquad \text{for} \quad \pi/4 \leqq \theta \leqq \pi/2 \end{aligned} \tag{2.18}$$

Finally substituting $\cot \theta = x/y$ we get at once from (2.18) that if the random variable $x$ has a bounded range $(-a, +a)$ the form of the probability density function of $w = x/y$ in the range $(0 \leqq w \leqq 1)$ is different from that in the range $(1 \leqq w \leqq \infty)$. The contradiction thus obtained leads to the proof of (3).

For the proof of (4) we introduce as usual the polar transformation $x = r \cos \theta$ and $y = r \sin \theta$ and integrate (2.17) with respect to $r$ over the range $(0, \infty)$. We further note that $\theta = \text{arc cot } x/y$ has a uniform distribution. Thus the equation for the probability density function of $\theta$ is given by

$$\int_0^\infty f(r \cos \theta)f(r \sin \theta)r\,dr = c_0 \tag{2.19}$$

where $c_0$ is a constant. Then substituting $x = r \cos \theta$ and $u = \tan \theta$ in (2.19) we get (2.16). Thus the problem of determining the entire class of distribution laws $F(x)$ is equivalent to that of complete enumeration of the solutions of the integral equation (2.16). This problem is very difficult and still remains to be solved.

**3. A characterization of the normal law.** We shall now derive a characterization of the normal distribution under some additional conditions on the distribution function $F(x)$. For this purpose, we give first some analytical lemmas which are also of independent interest.

LEMMA 3.1. *Let $\Phi(z)$ be a decomposable characteristic function which is regular (analytic) in a strip $-\alpha < \operatorname{Im} z < +\alpha$ $(\alpha > 0)$ of the complex z-plane. Let $\Phi_1(z)$ be a factor of $\Phi(z)$. Then the characteristic function $\Phi_1(z)$ is also regular at least in the strip $-\alpha < \operatorname{Im} z < +\alpha$.*

This lemma on the factorization of analytic characteristic functions is due to Raikov [5]. A proof of this lemma is presented by Loève ([2], p. 213).

LEMMA 3.2. *Let $\Phi(z)$ be a decomposable regular (analytic) characteristic function and $\Phi_1(z)$ a factor of $\Phi(z)$. Let $\Phi(-iv)$ exist for some $v$ ($v \neq 0$ real). Then for this $v$, $\Phi_1(-iv)$ must also exist. Further, there always exist two finite real numbers $K > 0$ and $a \geqq 0$ not depending on $v$ such that the inequality*

$$\Phi_1(-iv) \leqq Ke^{a|v|}\Phi(-iv) \tag{3.1}$$

*is satisfied.*

This lemma is also due to Raikov [5]. A proof of this lemma is presented by Loève ([2], p. 214).

LEMMA 3.3. *Under the same conditions as in Lemma 3.2, let $z = t + iv$ ($t$ and $v$ both real). Then we have the inequality*

$$| \Phi_1(-z) | \leqq Ke^{a|v|}\Phi(-iv). \tag{3.2}$$

The proof follows at once from (3.1) and the well-known property of the positive definite functions

$$\max_{-\infty \leqq t \leqq +\infty} | \Phi_1(t + iv) | \leqq \Phi_1(iv), \qquad (t \text{ and } v \text{ both real}).$$

LEMMA 3.4. *Let $f(x)$ be a continuous non-negative function of the real variable $x$. Let the integral $\int_0^\infty x^v f(x)\,dx$ exist for all real $v > 0$. Then the integral*

$$I(z) = \int_0^\infty x^{-iz} f(x)\,dx$$

*as a function of the complex variable $z$ is regular (analytic) in the upper half plane $\operatorname{Im} z > 0$. Conversely if the function $I(z)$ is regular in the upper half plane $\operatorname{Im} z > 0$, then the integral $\int_0^\infty x^v f(x)\,dx$ exists for all real $v > 0$.*

PROOF. We first note that $I(z)$ is uniformly convergent in every closed domain of the half plane $\operatorname{Im} z > 0$. Then using the well known theorems on regular functions ([6], pp. 107, 116) we derive that $I(z)$ is regular in the half plane $\operatorname{Im} z > 0$. The proof of the converse statement is obvious.

From Lemma 3.4, it is also easy to see that if the integral $\int_0^\infty x^v f(x)\,dx$ exists for all $v > 0$, then the integral $\int_0^\infty x^{iz} f(x)\,dx$, ($z$ complex) is regular in the lower half plane $\operatorname{Im} z < 0$.

LEMMA 3.5. *Under the same conditions as in Theorem 2.1, let the distribution law*

*$F(x)$ have finite moments of all orders. Let $\varphi(t) = E(e^{it\ln|x|})$ denote the characteristic function of the distribution of* $\ln|x|$. *Then $\varphi(z) = E(e^{iz\ln|x|})$ as a function of the complex variable $z$ is regular in the region* $\operatorname{Im} z < 1$.

PROOF. Since $F(x)$ has finite moments of all orders, the integral $\int_0^\infty x^v f(x)\,dx$ is convergent for all $v > 0$, where $f(x)$ is the probability density function. We further note that $f(x)$ is symmetric about the origin $x = 0$. Then applying Lemma 3.4, we get easily that

$$\varphi(z) = E(e^{iz\ln|x|}) = 2\int_0^\infty x^{iz} f(x)\,dx \tag{3.3}$$

is regular at least in the lower half plane $\operatorname{Im} z < 0$.

Next we note that the characteristic function $1/\cosh[(\pi/2)t]$ can be continued in the complex $z$-plane since $1/\cosh[(\pi/2)z]$ is regular in the strip $|\operatorname{Im} z| < 1$. Then applying Lemma 3.1 to the relation (2.14), we deduce at once that $\varphi(t)$ can also be continued in the complex $z$-plane and further $\varphi(z)$ is also regular at least in the strip $|\operatorname{Im} z| < 1$. Thus combining the two results we conclude that $\varphi(z)$ is regular in the region $\operatorname{Im} z < 1$. Similarly we see that $\varphi(-z)$ is regular in the region $\operatorname{Im} z > -1$.

We are now in a position to prove the following theorem.

THEOREM 3.1. *In addition to the conditions of Theorem 2.1, if the following two conditions are satisfied:*

(1) *$F(x)$ has finite moments of all orders,*

(2) *$\varphi(z) = E(e^{iz\ln|x|})$ has no zeros in its region of regularity ($z$ complex), then $F(x)$ is normal.*

We must note in this connection that the condition (2) is essential for the theorem. In the next section we shall give an example to show that the theorem is not true if the condition (2) is not satisfied.

PROOF. We examine more closely the equation

$$\varphi(z)\varphi(-z) = \frac{1}{\cosh\left(\frac{\pi}{2}z\right)} \tag{3.4}$$

for complex values of $z$.

For further investigation, we have to study the analytical behaviour of the function $\cosh[(\pi/2)z]$ in the complex $z$-plane. We note that $\cosh[(\pi/2)z]$ is an entire function of order unity having simple zeros at the points $z = \pm i(2k+1)$, $k = 0, 1, 2, \cdots$ on the imaginary axis. Then applying the decomposition theorem ([6], p. 299), we have the canonical representation of $\cosh[(\pi/2)z]$ as:

$$\cosh\left(\frac{\pi}{2}z\right) = \prod_{k=0}^{\infty}\left(1 + \frac{z^2}{\alpha_k^2}\right) \tag{3.5}$$

where $\alpha_k = 2k+1$; $k = 0, 1, 2, \cdots$. It is also easy to verify that the condition $\sum_{k=0}^{\infty} 1/\alpha_k^2 < \infty$ is satisfied.

From the conditions of the theorem 3.1 and lemma 3.5 it follows that the

characteristic function $\varphi(z)$ is regular in the region $\mathrm{Im}\, z < 1$ and has no zeros in this region. We now factorize $\varphi(z)$ in the following manner:

$$\varphi(z) = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1+iz}{2}\right) \theta(z). \tag{3.6}$$

From the elementary properties of the Gamma function ([6], p. 313) it can be verified easily that $\Gamma((1 + iz)/2)$ is a meromorphic function which is regular everywhere in the region $\mathrm{Im}\, z < 1$, real on the imaginary axis and has no zeros in its region of regularity. We also note that its reciprocal $1/\Gamma((1 + iz)/2)$ is an entire function of order unity having simple zeros at the points $z = i(2k + 1)$, $k = 0, 1, 2, \cdots$ all located on the imaginary axis ([6], p. 415). Hence using the factorization theorem of Hadamard ([6], p. 332) we get

$$\frac{\sqrt{\pi}}{\Gamma\left(\frac{1+iz}{2}\right)} = e^{i\rho z} \prod_{k=0}^{\infty} \left(1 - \frac{z}{i\alpha_k}\right) e^{z/i\alpha_k} \tag{3.7}$$

where $\rho \neq 0$ real; $\alpha_k = 2k + 1$, $k = 0, 1, 2, \cdots$. Thus the function $\theta(z)$ introduced in (3.6) must also be regular at least in the region $\mathrm{Im}\, z < 1$, real on the imaginary axis and without any zeros. From (3.6) we get

$$\varphi(z)\varphi(-z) = \frac{1}{\pi} \cdot \Gamma\left(\frac{1+iz}{2}\right) \cdot \Gamma\left(\frac{1-iz}{2}\right) \cdot \theta(z) \cdot \theta(-z). \tag{3.8}$$

Again it is easy to verify from (3.5) and (3.7)

$$\Gamma\left(\frac{1+iz}{2}\right) \cdot \Gamma\left(\frac{1-iz}{2}\right) = \frac{\pi}{\cosh\left(\frac{\pi}{2} z\right)}. \tag{3.9}$$

Hence using (3.8) and (3.9) we get easily from (3.4) that

$$\theta(z)\theta(-z) = 1 \tag{3.10}$$

holding for complex values of $z$. But we note that $\theta(-z)$ is regular at least in the region $\mathrm{Im}\, z > -1$ and has no zeros in its region of regularity. Hence $1/\theta(-z)$ is also regular at least in the region $\mathrm{Im}\, z > -1$ and without any zeros in this region. Then using the relation (3.10) it follows easily that $\theta(z)$ is regular everywhere throughout the complex plane, that is, it is an entire function. We note further that $\theta(z)$ has no zeros in the complex plane.

We next prove that the order of the entire function $\theta(z)$ cannot exceed unity. We apply the inequality (3.2) to the relation (3.4) and using the expression for $\varphi(z)$ in (3.6), we get after a little rearrangement

$$|\theta(z)| \cos\left(\frac{\pi}{2} v\right) \leqq K\sqrt{\pi} \cdot \frac{e^{a|v|}}{\left|\Gamma\left(\frac{1+iz}{2}\right)\right|} \leqq K\sqrt{\pi} \cdot \frac{e^{a|z|}}{\left|\Gamma\left(\frac{1+iz}{2}\right)\right|} \tag{3.11}$$

where $z = t + iv$ ($t$ and $v$ both real).

But the right hand side of (3.11) is an entire function of order unity. Hence from (3.11) it follows that the order of the entire function $\theta(z)$ cannot exceed unity. Again we have already proved that $\theta(z)$ has no zeros throughout the complex plane. Hence using the factorization theorem of Hadamard, we get $\theta(z) = e^{cz}$. Since $\theta(z)$ is real on the imaginary axis we get $\theta(z) = e^{i\alpha z}$ where $\alpha$ is real. Thus we obtain from (3.6)

$$\varphi(-z) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1-iz}{2}\right)e^{-i\alpha z}. \tag{3.12}$$

Next we substitute $z = iv$ ($v > 0$ real) in (3.12) and get

$$\varphi(-iv) = 2\int_0^\infty x^v f(x)\,dx = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1+v}{2}\right)e^{\alpha v}. \tag{3.13}$$

Since the distribution of $x$ is symmetric about the origin, all the moments of odd order are equal to zero and a moment of the even order $2k$ is given by

$$\mu_{2k} = \int_{-\infty}^\infty x^{2k} f(x)\,dx = 2\int_0^\infty x^{2k} f(x)\,dx. \tag{3.14}$$

Finally substituting $v = 2k$ ($k$ a positive integer) in (3.13) we have

$$\mu_{2k} = \frac{1}{\sqrt{\pi}}\Gamma(k+\tfrac{1}{2})e^{2\alpha k} = \frac{(2k)!}{k!2^k}\sigma^{2k} \tag{3.15}$$

where $\sigma = e^\alpha/\sqrt{2}$.

The proof of theorem (3.1) follows at once from the fact that the moments in (3.15) determine uniquely the normal distribution with mean zero and variance $\sigma^2$.

**4. An example.** The non-normal distribution functions constructed in [1], [7] have moments only up to a certain finite order. Here we give an example of a non-normal distribution having finite moments of all orders. We shall now construct a characteristic function $\varphi(z)$ which satisfies the basic equation (3.4), is regular in the region $\operatorname{Im} z < 1$, but having zeros in its region of regularity so that the condition (2) of Theorem 3.1 is violated. We give first two lemmas.

LEMMA 4.1. *Let*

$$\Phi(t) = \frac{\left(1+\frac{it}{\gamma}\right)\left(1+\frac{it}{\bar\gamma}\right)}{\left(1-\frac{it}{\alpha}\right)\left(1-\frac{it}{\gamma}\right)\left(1-\frac{it}{\bar\gamma}\right)} \tag{4.1}$$

*where* $\gamma = \alpha + i\beta$; $\bar\gamma = \alpha - i\beta$ *and* $\alpha > 0$, $\beta > 0$ *both real. Then* $\Phi(t)$ *is always a characteristic function whenever the relation* $\beta \geqq 2\sqrt{2}\,\alpha$ *is satisfied.* The proof follows from a more general result on rational characteristic functions ([4], p. 721).

LEMMA 4.2. *Let* $Q(z)$ *be an entire function of order unity having only purely imaginary zeros. Then its reciprocal* $1/Q(z)$ *is always a characteristic function.*

The proof follows from the result ([3], p. 140).

Next we define the quantities

$$
(4.2)\quad
\begin{aligned}
\alpha_k &= 2k + 1 & k &= 0, 1, 2, \cdots N, N + 1 \cdots \infty\\
\beta_k &\geqq 2\sqrt{2}\,\alpha_k & k &= 0, 1, 2, \cdots N \quad (N > 0)\\
\gamma_k &= \alpha_k + i\beta_k & k &= 0, 1, 2, \cdots N\\
\bar{\gamma}_k &= \alpha_k - i\beta_k & k &= 0, 1, 2, \cdots N
\end{aligned}
$$

and construct the function $\varphi(z)$ as:

$$
(4.3)\quad \varphi(z) = \prod_{k=0}^{N} \frac{\left(1 + \frac{z}{i\gamma_k}\right)\left(1 + \frac{z}{i\bar{\gamma}_k}\right)}{\left(1 - \frac{z}{i\alpha_k}\right)\left(1 - \frac{z}{i\gamma_k}\right)\left(1 - \frac{z}{i\bar{\gamma}_k}\right)e^{z/i\alpha_k}} \prod_{k=N+1}^{\infty} \frac{1}{\left(1 - \frac{z}{i\alpha_k}\right)e^{z/i\alpha_k}}
$$

$$
= P_1(z)\cdot P_2(z).
$$

From Lemma 4.1 it follows that $P_1(z)$ is a characteristic function, while we get as an immediate consequence of Lemma 4.2 that $P_2(z)$ is also a characteristic function. Hence $\varphi(z)$ in (4.3) is a characteristic function. It is also easy to verify that $\varphi(z)$ is regular in the region $\operatorname{Im} z < 1$ and has simple zeros at the points $z = -i\alpha_k \pm \beta_k$ $(k = 0, 1, 2, \cdots N)$ inside the region where $\alpha_k$ and $\beta_k$ are defined in (4.2). We also see easily that $\varphi(z)$ satisfies the basic equation (3.4). Then we take $\varphi(z)$ in (4.3) as the characteristic function of the distribution of $\ln |x|$ and verify at once that the corresponding distribution function $F(x)$ has moments of all orders, but is not normal and the quotient $x/y$ follows the Cauchy law.

In conclusion the author wishes to express his thanks to Professor Eugene Lukacs for some helpful comments.

## REFERENCES

[1] R. G. Laha, "An example of a non normal distribution where the quotient follows the Cauchy law," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 44 (1958), pp. 222–223.

[2] M. Loève, *Probability Theory*, D. Van Nostrand, New York, 1955.

[3] E. Lukacs and O. Szász, "Certain Fourier transforms of distributions," *Can. Jour. Math.*, Vol. 3 (1951), pp. 140–144.

[4] E. Lukacs, "Remarks concerning characteristic functions," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 717–723.

[5] D. A. Raikov, "On the decomposition of Gauss and Poisson laws," *Izv. Akad. Nauk. SSSR. Ser. Mat.*, Vol. 2 (1938), pp. 91–124 (in Russian).

[6] S. Saks and A. Zygmund, *Analytic Functions*, Warszawa and Wroclaw, 1952.

[7] G. P. Steck, "A uniqueness property not enjoyed by the normal distribution," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 604–606.

# CONTINUOUS SAMPLING PROCEDURES WITHOUT CONTROL[1]

By C. Derman, M. V. Johns, Jr., and G. J. Lieberman

*Columbia University, Stanford University, and Stanford University*

**1. Summary.** Several modifications of the Dodge CSP-1 procedure [1] are presented. Changes are made in the rule of action when a defective item is observed while on sampling. The Average Outgoing Quality Limit (AOQL) for these new procedures are derived without the assumption of control. These results are compared with the AOQL assuming control. A production process is said to be in statistical control if there is a constant probability $p$ that an item is defective, and if the states of all the items (defective or nondefective) are stochastically independent. Further, the AOQL for the CSP-1 procedure using probability sampling (looking at every item with probability $1/k$ when on sampling) is derived without the assumption of control.

**2. Introduction and results.** Two continuous sampling procedures are considered. The first procedure is denoted by CSP-4[2] and is as follows:

a) At the outset, inspect 100 per cent of the units consecutively as produced and continue such inspection until $i$ units in succession are found clear of defects.

b) When $i$ units in succession are found clear of defects, discontinue 100 per cent inspection, and inspect only a fraction $1/k$ of the units, choosing the item to be observed at random from a segment of size $k$ (this type of sampling will be called random sampling).

c) If a sample unit is found defective revert immediately to 100 per cent inspection, eliminating from the production process the remaining $(k - 1)$ items in the segment, and commencing 100 per cent inspection with the next item following the eliminated segment. Continue 100 per cent inspection until again $i$ units in succession are found clear of defects, as in paragraph (a).

d) Correct or replace with good units all defective units found.

It is important to discuss the implications of (c). These eliminated units can be considered as a source of good items for (d). Furthermore, under certain mathematical models for the production process such as "a state of statistical control" condition c is equivalent to the following:

> If a sample unit is found defective revert immediately to 100 per cent inspection, commencing such inspection with the segment in which the defective item is observed. Continue 100 per cent inspection until again $i$ units in succession are found clear of defects, as in paragraph (a).

The second continuous sampling procedure considered will be denoted by CSP-5 and is the same as CSP-4 except for condition (c) which is as follows:

---

Received October 20, 1958; revised June 18, 1959.

[1] This work was sponsored by the Office of Naval Research under contract N6onr-25126.

[2] CSP-2 and CSP-3 have already been used to denote other continuous sampling procedures.

c′) If a sample unit is found defective screen the remaining $k - 1$ items in the segment. Upon completion of this screening, commence 100 per cent inspection with the next item produced. Continue 100 per cent inspection until again $i$ units in succession, not including the $k - 1$ screened items, are found clear of defects, as in paragraph (a).

These procedures differ from the Dodge CSP-1 procedure in paragraphs (b) and (c). Dodge's statements [1] analogous to (b) and (c) are as follows:

When $i$ units in succession are found clear of defects, discontinue 100% inspection and inspect only a fraction $1/k$ of the units, selecting individual sample units one at a time from the flow of product, in such a manner as to assure an unbiased sample.

If a sample unit is found defective, revert immediately to a 100% inspection of succeeding units and continue until again $i$ units in succession are found clear of defects, as in paragraph (a).

It is not immediately evident what Dodge meant by the phrase, "· · · , *selecting individual sample units one at a time from the flow of product, in such a manner as to assure an unbiased sample.*" However, Dodge did study properties of his procedure and presented equations and charts for determining the Average Outgoing Quality Limit (AOQL) as functions of the parameters $k$ and $i$, under the assumption that the process is in a state of statistical control. There are several interpretations of the sampling procedure while on partial inspection which lead to Dodge's operating characteristics under the assumption of control. These are as follows: (1) look at every $k$th item. This type of sampling is denoted as systematic sampling and has the practical disadvantage that the particular item to be chosen is known in advance. (2) sample every item with probability $1/k$. This type of sampling is denoted as probability sampling and has the disadvantage that the number of uninspected items is a random variable. The result showing the coincidence of the operating characteristic using this type of sampling with CSP-1 is due to Resnikoff [2]. (3) sample only a fraction $1/k$ of the units, choosing the item to be observed at random from a segment of size $k$ (random sampling). If the sample unit is found defective begin 100% inspection with the item *following the segment* in which the defective item was observed, allowing the $k - 1$ uninspected items to enter into the production stream.

The CSP-4 and CSP-5 procedures are variations of this last type of sampling, i.e., random sampling. These procedures are investigated under the assumption of the existence of a state of statistical control and the AOQL's so obtained do not coincide exactly with the values given by Dodge for CSP-1. More important, however, the CSP-4 and CSP-5 procedures are analyzed without the assumption of the existence of a state of statistical control.

The problem of determining an AOQL for a Dodge type procedure without the assumption that the process is in a state of statistical control was first considered by Lieberman in [3], where it was shown that the CSP-1 procedure guarantees an AOQL whether or not the process is in a state of statistical control. In fact, for this case the AOQL equals $(k - 1)/(k + i)$. This result was obtained

under the hypothesis of random sampling while on partial inspection. The same result is obtained in this paper under the hypothesis of probability sampling while on partial inspection. For a given $k$ and $i$, the above value of the AOQL is always higher than that obtained using Dodge's equations. This is to be expected since the AOQL, without the assumption of control, is the least upper bound of the average quality level that a production process is able to achieve. This is not to imply that this is the average outgoing quality of a typical production process, but rather, that the average outgoing quality of the process can never exceed this AOQL value. The production process that actually achieves this level is one which alternates between producing all defective items during partial inspection and producing all non-defective items during 100% inspection.

It is the authors' contention that the assumption of control is not always justified. Whereas a production process which achieves the AOQL found by Lieberman seems unlikely, it should be emphasized that deviations from control can produce values of the average outgoing quality ranging up to the AOQL found by Lieberman.

It is intuitively clear that under CSP-4 and CSP-5 a production process which alternates between producing all defective items during partial inspection and producing all non-defective items during 100% inspection, will *not* represent the least favorable case. It is shown in this paper that both of these procedures guarantee a non-trivial AOQL whether or not the process is in a state of statistical control. In fact, for CSP-4

$$\text{AOQL} = \begin{cases} \dfrac{(c_4 + 2) - 2\sqrt{c_4 + 1}}{c_4^2}, & c_4 \neq 0 \\ \frac{1}{4}, & c_4 = 0 \end{cases} \quad \text{where } c_4 = (i - k + 1)/k$$

The AOQL is actually achieved when the process alternates between producing

$$d_4 = \begin{cases} \dfrac{k^2\sqrt{(i+1)/k} - k^2}{i - k + 1}, & i \neq k - 1 \\ k/2, & i = k - 1 \end{cases}$$

defective items in a block of size $k$ during partial inspection and producing all non-defective items during 100 per cent inspection. Similarly, for CSP-5

$$\text{AOQL} = \frac{(c_5 + 2) - 2\sqrt{c_5 + 1}}{c_5^2}, \quad \text{where } c_5 = i/k.$$

Note that the AOQL depends only on the ratio $i/k$, and not on the individual values. This AOQL is achieved when the process alternates between producing

$$d_5 = \frac{k^2\sqrt{i/k + 1} - k^2}{i}$$

defective items in a block of size $k$ during partial inspection and producing all non-defective items during 100 per cent inspection.

Naturally, these results are always higher than those obtained assuming control. However, the values of $d$ given are not so high as to be unrealistic. For example, if an operator knows that only 1 in $k$ items is to be chosen at random and observed, he may be careless enough to produce $d$ defective items in this block, whereas if he knows every item is to be looked at (100 per cent inspection) he will be very careful and produce all good items. Hence, the AOQL values given above may not be unreasonably large.

Finally, the CSP-4 or CSP-5 procedures are used in practice because of a reluctance to pass a segment in which a defective item has already been observed. Usually, the equations for the AOQL of CSP-1 under the assumption of control are used to find the necessary parameters $i$ and $k$ for the CSP-4 or CSP-5 procedures since this is a "conservative" approximation. However, its conservatism depends upon the realism of the assumption of control. It is interesting to point out that the CSP-5 procedure guarantees that the AOQL will never exceed 25% regardless of the choice of $i$ and $k$.

**3. Theorems and proof for the** AOQL **without the assumption of control for** CSP-4 **and** CSP-5. Define

$D_{st}$ = number of defects produced in the $s$th block of the $t$th cycle,

$$D_{st} = 0, 1, \cdots, k \text{ for all } s, t.$$

A cycle is the period where partial inspection begins to the time a defective is observed. A block is a segment of $k$ items produced while on partial inspection from which a single item is chosen at random for inspection.

$N_t$ = number of blocks (of $k$ items) sampled in the $t$th cycle. It is pointed out that the cycle terminates when a defective is found and that for the procedures considered the block in which the defective is drawn is not put directly into the production stream. However, it will still be considered as part of the $t$th cycle. Under CSP-4, the block is eliminated and under CSP-5, the block is screened replacing all defective items by good ones.

$X_t$ = total number of defects being passed in the $t$th cycle. $X_t = \sum_{s=1}^{N_t-1} D_{st}$.

$\delta_{st}$ are zero-one random variables and indicate whether the $s$th item in the 100% inspection sequence preceding the $t$th cycle of partial inspection are non-defective or defective.

$M_t$ = number of items inspected in the 100% inspection sequence preceding the $t$th cycle of partial inspection. This is a sure function of $\delta_{st}$.

A strategy of nature is characterized by a pair of doubly infinite sequences of possibly dependent random variables

$$\{\{D_{st}\}, \{\delta_{st}\}\}$$

Define the number $L_j$, $(j = 4, 5)$, as the smallest numbers with the property that for every process the probability is zero that

$$\limsup_{m\to\infty} \frac{\sum_{t=1}^{m} X_t}{k\sum_{t=1}^{m} N_t - m\alpha_j + \sum_{t=1}^{m} M_t} > L_j; \qquad (j = 4, 5) \tag{1}$$

where

$$\alpha_j = \begin{cases} k - 1, & j = 4 \\ 0, & j = 5 \end{cases}$$

The numbers $L_4$ and $L_5$ are called the AOQL for CSP-4 and CSP-5 respectively. It is evident that the ratio whose lim sup is taken in (1) is just the total number of defectives contributed to the outgoing product in the first $m$ cycles divided by the total number of items contributed to the outgoing product in the $m$ cycles.

It is clear that in order to determine $L$ we may confine ourselves to consideration of strategies of nature for which the number of cycles is infinite with probability 1. Furthermore, if we choose $\{\delta_{st}\} = \{0, 0, \cdots, 0, \cdots\}$ with probability 1, independent of the past, we are assured that $M_t = i$, $(t = 1, 2, \cdots)$, with probability 1. Hence, any strategy of nature for which the $\delta_{st}$ are not of this form is dominated by a corresponding strategy for which they are. Similarly it is sufficient to consider the special class of strategies for which the number of defectives in every block on partial inspection is $\geqq 1$. Hence, by confining ourselves to such strategies we may characterize nature's strategy by the single infinite sequence $\{D_{st}\}$, where the random variables $D_{st}$ take on the values $1, 2, \cdots, k$, with probability 1. It then follows that

$$\limsup_{m\to\infty} \frac{\sum_{t=1}^{m} X_t}{k\sum_{t=1}^{m} N_t - m\alpha_j + \sum_{t=1}^{m} M_t} \leqq \limsup_{m\to\infty} \frac{\sum_{t=1}^{m} X_t}{k\sum_{t=1}^{m} N_t - m\alpha_j + mi}; \tag{2}$$

$$(j = 4, 5).$$

THEOREM 1:[3] *For every strategy* $\{D_{st}\}$ *of nature and for all* $m$

$$\frac{\sum_{t=1}^{m} E(X_t \mid \underline{D}_t)}{k\sum_{t=1}^{m} E(N_t \mid \underline{D}_t) + m(i - \alpha_j)} \leqq L(c_j) \qquad (j = 4, 5); \tag{3}$$

[3] The authors are indebted to Professor S. Karlin for suggesting the method of proof used in this theorem.

*where*

$$L(c_j) = \begin{cases} \dfrac{(c_j + 2) - 2\sqrt{c_j + 1}}{c_j^2}, & c_j \neq 0 \\ \frac{1}{4}, & c_j = 0 \end{cases} \qquad (j = 4, 5); \tag{4}$$

$$c_j = \frac{i - \alpha_j}{k}; \tag{5}$$

*and*

$$\underline{D}_t = \{D_{1t}, D_{2t}, \cdots\}. \tag{6}$$

PROOF: We may write

$$X_t = \sum_{s=1}^{\infty} D_{st} U_{st}, \quad \text{where } U_{st} = \begin{cases} 1, N_t > s \\ 0, \text{otherwise} \end{cases} \tag{7}$$

Hence,

$$\begin{aligned} E(X_t \mid \underline{D}_t) &= \sum_{s=1}^{\infty} D_{st} E(U_{st} \mid \underline{D}_t) \\ &= D_{1t}\left(1 - \frac{D_{1t}}{k}\right) + D_{2t}\left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) \\ &\quad + D_{3t}\left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right)\left(1 - \frac{D_{3t}}{k}\right) + \cdots. \end{aligned} \tag{8}$$

This is a geometric series that is bounded uniformly by the convergent series $k\sum_{s=1}^{\infty} (1 - 1/k)^s$.

Similarly,

$$N_t = 1 + \sum_{s=1}^{\infty} U_{st}, \tag{9}$$

so that

$$\begin{aligned} E(N_t \mid \underline{D}_t) &= 1 + \left(1 - \frac{D_{1t}}{k}\right) \\ &+ \left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) + \left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right)\left(1 - \frac{D_{3t}}{k}\right) + \cdots. \end{aligned} \tag{10}$$

Again, this is uniformly bounded by a convergent geometric series.

From (8) it follows that

$$\sum_{t=1}^{m} E(X_t \mid \underline{D}_t) = \sum_{t=1}^{m} \left[ D_{1t}\left(1 - \frac{D_{1t}}{k}\right) + D_{2t}\left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) + \cdots \right]$$

$$\begin{aligned} &= \sum_{t=1}^{m} \left[ \left( \frac{D_{1t}\left(1 - \frac{D_{1t}}{k}\right)}{k + (i - \alpha_j)\frac{D_{1t}}{k}} \right) \left(k + (i - \alpha_j)\frac{D_{1t}}{k}\right) \right. \\ &\quad + \left( \frac{D_{2t}\left(1 - \frac{D_{2t}}{k}\right)}{k + (i - \alpha_j)\frac{D_{2t}}{k}} \right) \left(k + (i - \alpha_j)\frac{D_{2t}}{k}\right)\left(1 - \frac{D_{1t}}{k}\right) \\ &\quad + \left( \frac{D_{3t}\left(1 - \frac{D_{3t}}{k}\right)}{k + (i - \alpha_j)\frac{D_{3t}}{k}} \right) \left(k + (i - \alpha_j)\frac{D_{3t}}{k}\right) \\ &\quad \left. \cdot \left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) + \cdots \right]. \end{aligned} \tag{11}$$

From (10) it follows that

$$\begin{aligned} &k \sum_{t=1}^{m} E(N_t \mid \underline{D}_t) + m(i - \alpha_j) \\ &= \sum_{t=1}^{m} \left[ k + k\left(1 - \frac{D_{1t}}{k}\right) + k\left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) + \cdots + (i - \alpha_j) \right]. \end{aligned} \tag{12}$$

Noting that

$$\frac{D_{1t}}{k} + \frac{D_{2t}}{k}\left(1 - \frac{D_{1t}}{k}\right) + \frac{D_{3t}}{k}\left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) + \cdots = 1$$

since the left hand side is just the probability of ultimately achieving a success when performing successive Bernoulli trials with success probabilities bounded away from zero, we see that expression (12) can be written as

$$\begin{aligned} &k \sum_{t=1}^{m} E(N_t \mid \underline{D}_t) + m(i - \alpha_j) \\ &= \sum_{t=1}^{m} \left[ \left(k + (i - \alpha_j)\frac{D_{1t}}{k}\right) + \left(k + (i - \alpha_j)\frac{D_{2t}}{k}\right)\left(1 - \frac{D_{1t}}{k}\right) \right. \\ &\quad \left. + \left(k + (i - \alpha_j)\frac{D_{3t}}{k}\right)\left(1 - \frac{D_{1t}}{k}\right)\left(1 - \frac{D_{2t}}{k}\right) + \cdots \right]. \end{aligned} \tag{13}$$

Hence,

$$\frac{\sum_{t=1}^{m} E(X_t \mid \underset{\sim}{D}_t)}{k \sum_{t=1}^{m} E(N_t \mid \underset{\sim}{D}_t) + m(i - \alpha_j)}$$

is merely a non-negatively weighted average of quantities of the form

$$f(D_{st}; i, k) = \frac{D_{st}\left(1 - \frac{D_{st}}{k}\right)}{k + (i - \alpha_j)\frac{D_{st}}{k}}, \qquad (j = 4, 5; s = 1, 2, \cdots) \tag{14}$$

and has an upper bound obtained by maximizing each of these expressions independently. Taking the derivative of (14) with respect to the value $d_{st}$ of $D_{st}$ (treated as a continuous variable) we obtain

$$f'(d_{st}, i, k) = \begin{cases} \dfrac{k - 2d_{st}}{k^2 + (i - \alpha_j)d_{st}} - \dfrac{(kd_{st} - d_{st}^2)(i - \alpha_j)}{[k^2 + (i - \alpha_j)d_{st}]^2}, & i \neq \alpha_j \\[2ex] \dfrac{1}{k} - 2\dfrac{d_{st}}{k^2}, & i = \alpha_j \end{cases} \tag{15}$$

The quantity $f(d_{st}, i, k)$ is clearly maximized by setting (15) equal to zero. Denoting the maximizing value of $d_{st}$ by $d_j$ since it is independent of $s$ and $t$ we obtain

$$d_j = \begin{cases} \dfrac{k^2\sqrt{(i - \alpha_j)/k + 1} - k^2}{(i - \alpha_j)}, & i \neq \alpha_j \\[2ex] k/2, & i = \alpha_j \end{cases} \qquad (j = 4, 5); \tag{16}$$

It then follows that

$$\begin{aligned} \frac{\sum_{t=1}^{m} E(X_t \mid \underset{\sim}{D}_t)}{k \sum_{t=1}^{m} E(N_t \mid \underset{\sim}{D}_t) + m(i - \alpha_j)} &\leq \frac{d_j\left(1 - \frac{d_j}{k}\right)}{k + (i - \alpha_j)\frac{d_j}{k}} \\ &= \left.\begin{cases} \dfrac{(c_j + 2) - 2\sqrt{c_j + 1}}{c_j^2}, & c_j \neq 0 \\[2ex] 1/4, & c_j = 0 \end{cases}\right\} = L(c_j); \qquad (j = 4, 5); \end{aligned} \tag{17}$$

where $c_j = (i - \alpha_j)/k$.

THEOREM 2: *For any strategy* $\{D_{st}\}$ *of nature, for either* CSP-4 *or* CSP-5

$$\lim_{m \to \infty} \left[\frac{1}{m} \sum_{t=1}^{m} X_t - \frac{1}{m} \sum_{t=1}^{m} E(X_t \mid \underset{\sim}{D}_t)\right] = 0, \tag{18}$$

*with probability* 1, *and*

$$\lim_{m\to\infty}\left[\frac{1}{m}\sum_{t=1}^{m} N_t - \frac{1}{m}\sum_{t=1}^{m} E(N_t \mid \underset{\sim}{D}_t)\right] = 0, \tag{19}$$

*with probability* 1.

PROOF: For $t = 1, 2, \cdots$, let

$$Z_t = X_t - E(X_t \mid \underset{\sim}{D}_t). \tag{20}$$

Then

$$E(Z_t \mid \underset{\sim}{D}_t) = E[X_t - E(X_t \mid \underset{\sim}{D}_t) \mid \underset{\sim}{D}_t] = 0 \tag{21}$$

so that $E(Z_t) = 0$. Furthermore, for $t > s$, $Z_t$ and $Z_s$ are conditionally independent given $\underset{\sim}{D}_t$ so that

$$E(Z_sZ_t) = E[E(Z_sZ_t \mid \underset{\sim}{D}_t)] = E[E(Z_t \mid \underset{\sim}{D}_t)\, E(Z_s \mid \underset{\sim}{D}_t)] = 0. \tag{22}$$

Now

$$\begin{aligned} E(Z_t^2) &= E(X_t^2) - E[E^2(X_t \mid \underset{\sim}{D}_t)] \leqq E(X_t^2) < k^2E(N_t^2) \\ &= k^2E[E(N_t^2 \mid \underset{\sim}{D}_t)] \leqq k^2 \sum_{s=1}^{\infty} s^2\left(1 - \frac{1}{k}\right)^{s-1} < \infty, \end{aligned} \tag{23}$$

since $D_{st} \geqq 1$ with probability 1. Now by a well known Law of Large Numbers for sums of orthogonal random variables ([4] Chapter IV, Theorem 5.2) equation (22) together with the uniform boundedness of $E(Z_t^2)$ shown by (23) implies that

$$\lim_{m\to\infty}\frac{1}{m}\sum_{t=1}^{\infty} Z_t = 0, \tag{24}$$

with probability 1,
so that (18) is established. Letting $Z_t^* = N_t - E(N_t \mid \underset{\sim}{D}_t)$, the proof of (19) is similar.

THEOREM 3. *For any strategy* $\{D_{st}\}$ *of nature*

$$L_j \leqq L(c_j) \qquad (j = 4, 5). \tag{25}$$

PROOF. By Theorem 1 we have

$$\frac{1}{m}\sum_{t=1}^{m} E(X_t \mid \underset{\sim}{D}_t) - L(c_j)\left[\frac{k}{m}\sum_{t=1}^{m} E(N_t \mid \underset{\sim}{D}_t) + (i - \alpha_j)\right] \leqq 0, \tag{26}$$

for all $m$. If for each $m$ we let

$$V_m = \frac{1}{m}\sum_{t=1}^{m} X_t - \frac{1}{m}\sum_{t=1}^{m} E(X_t \mid \underset{\sim}{D}_t), \tag{27}$$

and

$$V_m' = \frac{1}{m}\sum_{t=1}^{m} N_t - \frac{1}{m}\sum_{t=1}^{m} E(N_t \mid \underset{\sim}{D}_t), \tag{28}$$

then by Theorem 2, $\lim_{m\to\infty} V_m = \lim_{m\to\infty} V_m' = 0$ with probability 1. But from (26) we have

$$\frac{\sum_{t=1}^{m} X_t}{k \sum_{t=1}^{m} N_t + m(i - \alpha_j)} \leq L(c_j) + \frac{V_m - kL(c_j)V'_m}{\frac{k}{m} \sum_{t=1}^{m} N_t + (i - \alpha_j)}, \tag{29}$$

and (25) follows upon taking the lim $\sup_{m\to\infty}$ of both sides of (29).

If we now let

$$d_j^* = \text{integer nearest to} \begin{cases} \dfrac{k^2\sqrt{(i-\alpha_j)/k+1} - k^2}{i - \alpha_j}, & i \neq \alpha_j \\ k/2, & i = \alpha_j \end{cases} \qquad (j = 4, 5); \tag{30}$$

then we have

THEOREM 4: *If the production process alternates between producing $d_4^*$ $[d_5^*]$ defective items in blocks of size k during partial inspection and all non-defective items during* 100 *per cent inspection, then for* CSP-4 [CSP-5]

$$\limsup_{m\to\infty} \frac{\sum_{t=1}^{m} X_t}{k \sum_{t=1}^{m} N_t + m(i - \alpha_j)}$$

*equals $L_4(c)$ $[L_5(c)]$ (approximately, due to the discreteness of $d_4^*$ and $d_5^*$) and hence the* AOQL *is given by $L_4(c)$ $[L_5(c)]$.*

PROOF: This result follows immediately from (16) and Theorems 2 and 3.

We remark that it is easily verified by differentiation that $L(c_5) \leq \lim_{c\to 0} L(c_5) = 1/4$, so that the AOQL $\leq \frac{1}{4}$ for CSP-5 for any choice of $i$ and $k$. We further remark that if defective items found when on 100 per cent inspection are not replaced by good items but are discarded, the previously derived results are still applicable, i.e., the AOQL is still given approximately by $L(c_j)$. If, under the CSP-4 procedure, a unit found defective while on sampling is also discarded together with the remaining $(k - 1)$ items and not replaced, the previously derived results are also applicable provided that $\alpha_j$ is set equal to $k$.

**4. CSP-4 and CSP-5 under control.** This section will be devoted to determining the Average Outgoing Quality (AOQ) function and the AOQL for the CSP-4 and CSP-5 procedures under the assumption of the existence of a state of statistical control.

The AOQ function is defined as

$$\begin{aligned} \text{AOQ}_j &= \limsup_{m\to\infty} \frac{\sum_{t=1}^{m} X_t}{k \sum_{t=1}^{m} N_t - m\alpha_j + \sum_{t=1}^{m} M_t} \\ &= \limsup_{m\to\infty} \frac{\sum_{t=1}^{m} X_t/m}{k \sum_{t=1}^{m} N_t/m - \alpha_j + \sum_{t=1}^{m} M_t/m} \qquad (j = 4, 5) \end{aligned} \tag{31}$$

where

$$\alpha_j = \begin{cases} k - 1, & j = 4 \\ 0, & j = 5 \end{cases}$$

Under the assumption of the existence of a state of statistical control at level $p$, the law of large numbers becomes applicable so that the AOQ function can be expressed as

$$\text{AOQ}_j = \frac{E(X_i)}{kE(N_i) - \alpha_j + E(M_i)} \qquad (j = 4, 5). \tag{32}$$

It is easily verified that

$$E(M_i) = \frac{1 - q^i}{pq^i} \tag{33}$$

$$E(N) = \frac{1}{p} \tag{34}$$

and

$$E(X_i) = (k - 1)q \tag{35}$$

where $q = 1 - p$. Hence,

$$\text{AOQ}_4 = \frac{(k - 1)(q^{i+1} - q^{i+2})}{1 + q^{i+1}(k - 1)} = \frac{(k - 1)pq^{i+1}}{1 + (k - 1)q^{i+1}}. \tag{36}$$

The maximizing value of $q$ for a fixed $i$ and $k$ is given by solving for $q$ the expression

$$(k - 1)\, q^{i+2} + (i + 2)\, q = (i + 1). \tag{37}$$

Denote this value by $q_{\max\text{-}4}$. The AOQL can then be written as

$$\text{AOQL}_4 = 1 - q_{\max\text{-}4} \frac{(i + 2)}{(i + 1)} \tag{38}$$

or, solving for $q_{\max\text{-}4}$, the expression

$$q_{\max\text{-}4} = (1 - \text{AOQL}_4) \frac{(i + 1)}{(i + 2)} \tag{39}$$

is obtained. Substituting this expression for $q_{\max\text{-}4}$ into (37) and solving for $k$, the relationship between $k$ and $i$ for a fixed AOQL is obtained, i.e.,

$$k = 1 + \left(\frac{i + 2}{i + 1}\right)^{i+2} \frac{(i + 1)\,\text{AOQL}_4}{(1 - \text{AOQL}_4)^{i+2}}. \tag{40}$$

For fixed $k$ and $i$, the expression for the AOQL for the CSP-4 procedure assuming control never exceeds the AOQL which is obtained without making any assumptions about the behavior of the process. However, the differences are much smaller for this procedure than for the CSP-1 procedure.

Similarly, for CSP-5, the AOQ function can be written as

$$\text{(41)} \qquad \text{AOQ}_5 = \frac{[q^{i+1} - q^{i+2}](k-1)}{1 + q^i(k-1)} = \frac{(k-1)pq^{i+1}}{1 + (k-1)q^i}.$$

The maximizing value of $q$ for a fixed $i$ and $k$ is given by solving for $q$ the expression

$$\text{(42)} \qquad 2(k-1)q^{i+1} - (k-1)q^i + (i+2)q = i + 1.$$

Denote this value by $q_{\max\text{-}5}$. The AOQL can then be written as

$$\text{(43)} \qquad \text{AOQL}_5 = \frac{(i+1)q_{\max\text{-}5} - (i+2)q^2_{\max\text{-}5}}{i}$$

or, solving for $q_{\max\text{-}5}$, the expression

$$\text{(44)} \qquad q_{\max\text{-}5} = \frac{(i+1) + \sqrt{(i+1)^2 - 4i(i+2)\,\text{AOQL}_5}}{2(i+2)}$$

is obtained.

Substituting this expression for $q_{\max\text{-}5}$ into (42) and solving for $k$, the relationship between $k$ and $i$ for a fixed AOQL is obtained, i.e.,

$$\text{(45)} \qquad k = 1 + \frac{(i+1) - (i+2)q_{\max\text{-}5}}{2q^{i+1}_{\max\text{-}5} - q^i_{\max\text{-}5}}.$$

Curves of constant AOQL derived from expressions (40), (44), and (45) are given in Figure 1.

**5. CSP-1 without assuming control and using probability sampling.** In this section, CSP-1 will be studied without assuming control but using a sampling procedure such that while on partial inspection, every item will be inspected with probability $1/k$, or passed without inspection with probability $(1-1/k)$. The notation of Sections 2 and 3 will be used, but for this problem $k$ need not be an integer but may be any number $>1$.

If we let $N_t^*$ denote the number of items contributed to the production stream during the $t$th partial inspection cycle, then the AOQL is defined, as before, as the smallest number $L$ with the property that for every strategy of nature the probability is zero that

$$\text{(46)} \qquad \limsup_{m\to\infty} \frac{\sum_{t=1}^m X_t}{\sum_{t=1}^m N_t^* + \sum_{t=1}^m M_t} > L.$$

To obtain the AOQL it is again sufficient to consider the special class of strategies of nature such that $M_t = i$ for all $t$, and we must investigate the quantity

$$\text{(47)} \qquad \limsup_{m\to\infty} \frac{\sum_{t=1}^m X_t}{\sum_{t=1}^m N_t^* + mi},$$

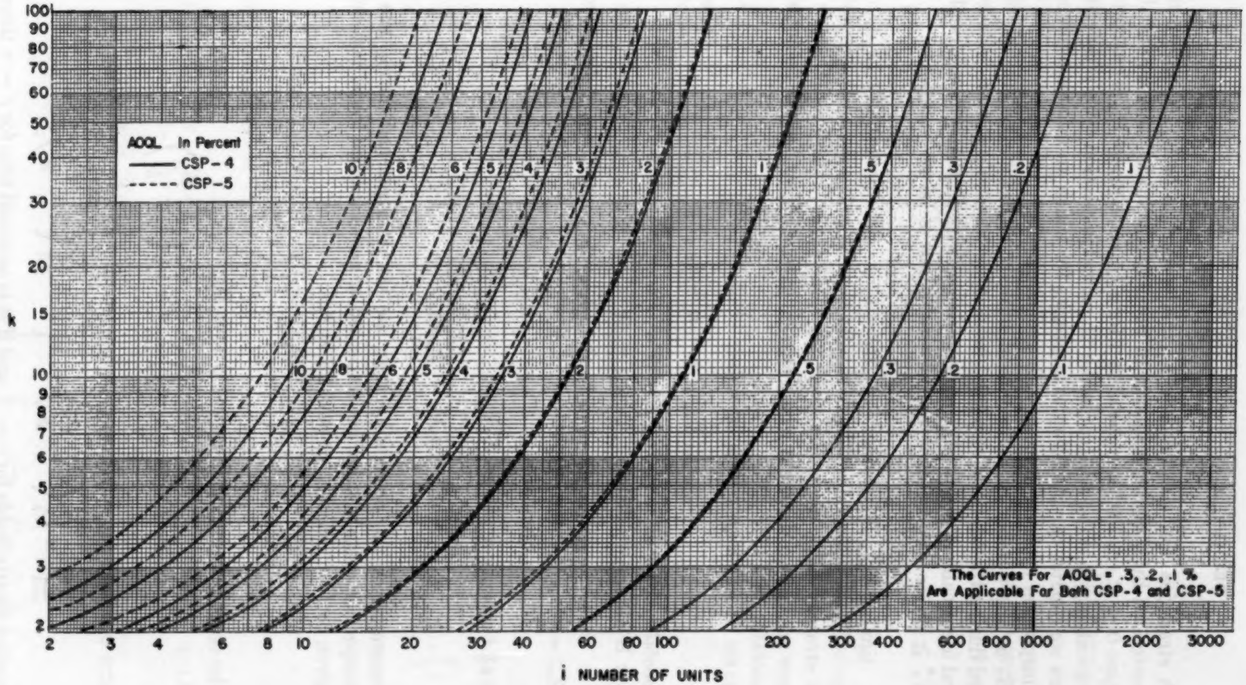


Fig. 1. Curves for Determining Values of $k$ and $i$ for A Given Value of AOQL for CSP-4 and CSP-5 under Control.

for such strategies. For this problem a (randomized) strategy of nature may be characterized by a double sequence of possibly dependent random variables $\{P_{st}\}$ where $0 \leqq P_{st} \leqq 1$ with probability 1 for all $s, t$ and where $P_{st}$ is interpreted as the probability that the $s$th item in the $t$th partial inspection cycle is defective. As before we restrict our attention to strategies for which an infinite number of partial inspection cycles will occur with probability 1.

Let $R_t$ be the number of items passed until (and including) the first item inspected during the $t$th cycle of partial inspection. Then the $R_t$'s are independently and identically distributed random variables with $E(R_t) = k$ and, furthermore, $N_t^* \geqq R_t$ for each $t$. Hence, by the Strong Law of Large Numbers

$$\liminf_{m\to\infty} \frac{1}{m} \sum_{t=1}^{m} N_t^* \geqq \lim_{m\to\infty} \frac{1}{m} \sum_{t=1}^{m} R_t = k, \quad \text{with probability 1,} \tag{48}$$

for any strategy $\{P_{st}\}$ of nature.

We now prove two theorems which enable us to characterize the behavior of the numerator of (47).

THEOREM 5: *For any strategy of nature* $\{P_{st}\}$

$$E(X_t \mid \underline{P}_t) = k - 1, \tag{49}$$

*with probability* 1 *for all* $t$, *where* $\underline{P}_t = \{P_{1t}, P_{2t}, \cdots\}$.

PROOF: If all $s, t$ we define

$$Z_{st} = \begin{cases} 1, & \text{if the } s\text{th item in the } t\text{th cycle contributes a defective to} \\ & \quad \text{the output,} \\ 0, & \text{otherwise,} \end{cases} \tag{50}$$

then for all $t$ we may represent $X_t$ by

$$X_t = \sum_{s=1}^{\infty} Z_{st}. \tag{51}$$

Furthermore, since the probability that the $s$th item reached during the $t$th partial inspection cycle is either not inspected or inspected and found non-defective is given by $(1 - P_{st}/k)$, we have for all $s, t$

$$E(Z_{st} \mid \underline{P}_t) = \left(1 - \frac{1}{k}\right) P_{st} \prod_{j=1}^{s-1} \left(1 - \frac{P_{jt}}{k}\right), \tag{52}$$

where the empty product is interpreted as 1. Hence,

$$E(X_t \mid \underline{P}_t) = \left(1 - \frac{1}{k}\right) \sum_{s=1}^{\infty} P_{st} \prod_{j=1}^{s-1} \left(1 - \frac{P_{jt}}{k}\right). \tag{53}$$

We now establish the following equation for all $r \geqq 1$ by induction:

$$\sum_{s=1}^{r} P_{st} \prod_{j=1}^{s-1} \left(1 - \frac{P_{jt}}{k}\right) = k\left[1 - \prod_{j=1}^{r} \left(1 - \frac{P_{jt}}{k}\right)\right]. \tag{54}$$

The equation clearly holds for $r = 1$, and if it is assumed true for $r = n$ then for $r = n + 1$ the left hand side becomes

$$(55)\quad k\left[1-\prod_{j=1}^{n}\left(1-\frac{P_{jt}}{k}\right)\right]+P_{n+1,t}\prod_{j=1}^{n}\left(1-\frac{P_{jt}}{k}\right) = k\left[1-\prod_{j=1}^{n+1}\left(1-\frac{P_{jt}}{k}\right)\right],$$

and the proof by induction is complete.

We now remark that if the number of partial inspection cycles occurring is to be infinite with probability 1, then we must have $\lim_{r\to\infty} P\{N_t^* > r\} = 0$ for each $t$, which implies that

$$(56)\quad \lim_{r\to\infty} P\{N_t^* > r \mid \underline{P}_t\} = \lim_{r\to\infty}\prod_{j=1}^{r}\left(1-\frac{P_{jt}}{k}\right) = 0,$$

with probability 1 for all strategies $\{P_{st}\}$ under consideration. The desired result (49) now follows from (53), (54) and (56).

THEOREM 6. *For any strategy of nature* $\{P_{st}\}$

$$(57)\quad E(X_t^2) \leqq 2(k-1)^2 + (k-1)$$

*for all* $t$.

PROOF: As in Theorem 5 we have

$$(58)\quad E(X_t^2 \mid \underline{P}_t) = 2\sum_{v=2}^{\infty}\sum_{w=1}^{v-1} E(Z_{vt}Z_{wt} \mid \underline{P}_t) + \sum_{s=1}^{\infty} E(Z_{st} \mid \underline{P}_t),$$

and for $v > w$

$$(59)\quad \begin{aligned} E(Z_{vt}Z_{wt} \mid \underline{P}_t) &= \left[\prod_{s=1}^{w-1}\left(1-\frac{P_{st}}{k}\right)\right]\left[P_{wt}\left(1-\frac{1}{k}\right)\right] \\ &\quad\cdot\left[\prod_{s=w+1}^{v-1}\left(1-\frac{P_{st}}{k}\right)\right]\left[P_{vt}\left(1-\frac{1}{k}\right)\right] \\ &= \left(1-\frac{1}{k}\right)^2 P_{vt}P_{wt}\prod_{\substack{s=1\\ s\neq w}}^{v-1}\left(1-\frac{P_{st}}{k}\right) \\ &= k\left(1-\frac{1}{k}\right)^2\frac{P_{vt}P_{wt}}{k-P_{wt}}\prod_{s=1}^{v-1}\left(1-\frac{P_{st}}{k}\right). \end{aligned}$$

Hence noting (49) of Theorem 5 we may write (58) as

$$(60)\quad E(X_t^2 \mid \underline{P}_t) = 2k\left(1-\frac{1}{k}\right)^2\sum_{v=2}^{\infty} P_{vt}\prod_{s=1}^{v-1}\left(1-\frac{P_{st}}{k}\right)\sum_{w=1}^{v-1}\frac{P_{wt}}{k-P_{wt}} + (k-1).$$

We now establish the following equation for all $r \geqq 2$ by induction:

$$(61)\quad \sum_{v=2}^{r} P_{vt}\prod_{s=1}^{v-1}\left(1-\frac{P_{st}}{k}\right)\sum_{w=1}^{v-1}\frac{P_{wt}}{k-P_{wt}} = k\left[1-\left(1+\sum_{w=1}^{r}\frac{P_{wt}}{k-P_{wt}}\right)\prod_{s=1}^{r}\left(1-\frac{P_{st}}{k}\right)\right].$$

It is easily verified that for $r = 2$ both sides of (61) are equal to $P_{1t} P_{2t}/k$. If (61) is assumed to be true for $r = n$ then for $r = n + 1$ the left hand side may written as

$$k\left[1 - \left(1 + \sum_{w=1}^{n} \frac{P_{wt}}{k - P_{wt}}\right)\prod_{s=1}^{n}\left(1 - \frac{P_{st}}{k}\right) + \frac{P_{n+1,t}}{k}\prod_{s=1}^{n}\left(1 - \frac{P_{st}}{k}\right)\sum_{w=1}^{n}\frac{P_{wt}}{k - P_{wt}}\right]$$

$$(62) \qquad = k\left[1 - \left(1 - \frac{P_{n+1,t}}{k}\right)\sum_{w=1}^{n}\frac{P_{wt}}{k + P_{wt}}\right)\prod_{s=1}^{n}\left(1 - \frac{P_{st}}{k}\right)\right]$$

$$= k\left[1 - \left(\frac{k}{k - P_{n+1,t}} + \sum_{w=1}^{n}\frac{P_{wt}}{k + P_{wt}}\right)\prod_{s=1}^{n+1}\left(1 - \frac{P_{st}}{k}\right)\right]$$

$$= k\left[1 - \left(1 + \sum_{w=1}^{n+1}\frac{P_{wt}}{k + P_{wt}}\right)\prod_{s=1}^{n+1}\left(1 - \frac{P_{st}}{k}\right)\right],$$

which is the right hand side of (61) with $r = n + 1$, so that the proof by induction is complete.

Now (60) and (61) imply that

$$(63) \qquad E(X_t^2 \mid \underline{P}_t) \leq 2k^2\left(1 - \frac{1}{k}\right)^2 + (k - 1) = 2(k - 1)^2 + (k - 1),$$

and the desired result (57) follows. An examination of (61) shows that if the $P_{st}$'s are (for example) bounded away from zero then equality holds in (57).

We now prove the main result of this section.

THEOREM 7. For CSP-1 with probability sampling the AOQL is given by

$$(64) \qquad L = \frac{k - 1}{k + i},$$

and this value of $L$ is achieved by (47) when nature's strategy is to produce all defective items during partial sampling and all non-defective items during 100% sampling.

PROOF: The results of Theorems 5 and 6 together with the argument used in Theorem 2 imply that

$$(65) \qquad \lim_{m\to\infty} \frac{1}{m}\sum_{t=1}^{m} X_t = k - 1, \qquad \text{with probability} \quad 1,$$

for any strategy $\{P_{st}\}$ of nature. This result together with (48) implies that

$$(66) \qquad L \leq \frac{k - 1}{k + i}.$$

The fact that equality holds in (66) follows by applying the Strong Law of Large Numbers to the quantities

$$\frac{1}{m}\sum_{t=1}^{m} N_t^* \quad \text{and} \quad \frac{1}{m}\sum_{t=1}^{m} X_t \tag{67}$$

for the case where nature uses the strategy described in the Theorem above.

## REFERENCES

[1] H. F. Dodge, "A sampling inspection plan for continuous production," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 264–279.
[2] George J. Resnikoff, "Some modifications of the Lieberman-Solomon multi-level continuous sampling plan, MLP" Technical Report No. 26, Applied Math. and Stat. Laboratory, Stanford University, Stanford California, 1956.
[3] Gerald J. Lieberman, "A note on Dodge's continuous inspection plan" *Ann. Math. Stat.*, Vol. 24 (1953), pp. 480–484.
[4] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, Inc., New York, 1953.

# SOME CONVERGENCE THEOREMS FOR STATIONARY STOCHASTIC PROCESSES[1]

By T. Kawata

*Tokyo Institute of Technology and Princeton University*

**1. Introduction.** Let $\mathcal{E}(t)$ $(-\infty < t < \infty)$ be a continuous stationary process of the second order (in the wide sense) with mean zero; that is,

$$E\mathcal{E}(t+u)\mathcal{E}(t) = \rho(u) \tag{1.1}$$

is a continuous function of $u$ only, and

$$E\mathcal{E}(t) = 0, \qquad -\infty < t < \infty. \tag{1.2}$$

Here $E$ means the expectation of a random variable.

We have, then,

$$\mathcal{E}(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dZ(\lambda), \tag{1.3}$$

and

$$\rho(u) = \int_{-\infty}^{\infty} e^{iu\lambda}\, dF(\lambda), \tag{1.4}$$

where $F(\lambda)$ is a bounded non-decreasing function such that

$$F(+\infty) - F(-\infty) = \rho(0) = E\,|\,\mathcal{E}(t)|^2,$$

and $Z(\lambda)$ is an orthogonal process such that

$$E\,|\,Z(\lambda') - Z(\lambda)|^2 = F(\lambda' - 0) - F(\lambda - 0). \tag{1.5}$$

$F(u)$ and $Z(\lambda)$ are called the spectral function and the random spectral function of $\mathcal{E}(t)$ respectively. (See, e.g., Doob [5], Chapter XI). Let

$$X(t) = f(t) + \mathcal{E}(t), \qquad -\infty < t < \infty, \tag{1.6}$$

where $f(t)$ is a numerical valued function, and consider

$$\int_{-\infty}^{\infty} x(t-s)K(s,n)\, ds, \tag{1.7}$$

$K(s, n)$ being also a numerical valued function depending on a parameter $n$.

Integrals of the type (1.7) appear in many fields in the theory of probability and statistics. For instance, we often encounter (1.6) in the problem of smoothing data of observed values, in the problem of predicting future values of $x(t)$, and in the problem of estimating the spectral density of a stationary process.

Received January 6, 1958; revised April 22, 1959.

[1] Work done while I was a research fellow at Princeton University in 1957–58, supported by The Rockefeller Foundation.

But here we shall consider the behavior of (1.6) from the analytical point of view along the lines of the classical theory of the general Fourier integral, and we shall show convergence theorems, some of which may be already known implicitly.

Next we shall consider the special case of (1.7)

$$\frac{1}{T}\int_{-T}^{T} X(s)K(s)\,ds = J(T). \tag{1.7a}$$

If $K(s) = e^{-i\xi s}$, then $|J(T)|^2$ may be considered as a function similar to the periodogram, in which $f(t)$ is a trigonometric polynomial and $\mathcal{E}(t) \equiv 0$. It is known that if $F(\lambda)$ is absolutely continuous and $p(\lambda) = F'(\lambda)$ (the spectral density of $\mathcal{E}(t)$), then $E\,|J(T)|^2$ converges to $p(\xi)$ provided $p(\xi)$ is continuous at $\xi$. We shall treat the convergence of $|J(T)|^2$ itself.

**2. Preliminaries.** Let the spectral function of the continuous (weakly) stationary process $\mathcal{E}(t)$ be $F(\lambda)$ as in the preceding section. Then the necessary and sufficient condition for the existence of

$$\eta(t) = \int_{-\infty}^{\infty} \mathcal{E}(t-s)\,dL(a)^{2} \tag{2.1}$$

for every $s$ is that there exists a function $k(x)$ such that

$$\int_{-\infty}^{\infty} |k(x)|^2\,dF(x) < \infty$$

and

$$\lim_{\substack{A\to-\infty \\ B\to\infty}} \int_{-\infty}^{\infty} \left| \int_{A}^{B} e^{ixs}\,dL(s) - k(x) \right|^2 dF(x) = 0,$$

where we assume that $L(s)$ is a function of bounded variation in every finite interval. $k(x)$ is called the Fourier-Stieltjes transform of $L(s)$ with respect to $F(x)$. In particular if $K(x)\ \varepsilon\ L_1(-\infty, \infty)$, then

$$\int_{-\infty}^{\infty} \mathcal{E}(t-s)K(s)\,ds \tag{2.2}$$

exists.

We frequently use the following lemmas which are very well known.

LEMMA 2.1:

(i) *The stochastic process* (2.1) *is also a stationary process in the wide sense and we have* $E\eta(t) = 0$ *and*

$$E\eta(t+u)\overline{\eta(t)} = \int_{-\infty}^{\infty} |k(x)|^2 \cdot e^{iux}\,dF(x), \tag{2.3}$$

*where* $F(x)$ *is the spectral function of* $\mathcal{E}(t)$.

---

[2] The integral is taken here as $\text{l.i.m.}_{B\to\infty}^{A\to-\infty} \int_A^B \mathcal{E}(t-s)dL(s)$, where l.i.m. means the limit in the mean of order 2 and the finite integral in the definition is also defined as a Riemann-Stieltjes integral, the limit process being taken as l.i.m. See M. Loève [10] or J. L. Doob [5].

(ii) *If we are given another process*

$$\eta_1(t) = \int_{-\infty}^{\infty} \mathcal{E}(t-s)\, dL_1(s), \tag{2.4}$$

*where $L_1(s)$ is of bounded variation in every finite interval and (2.4) is assumed to exist, then*

$$E\eta(t+u)\overline{\eta_1(t)} = \int_{-\infty}^{\infty} k(x)\overline{k_1(x)}e^{iux}\, dF(x), \tag{2.5}$$

*$k_1(x)$ being the Fourier-Stieltjes transform of $L_1(s)$ with respect to $F(x)$.*

LEMMA 2.2: *The stochastic process $\eta(t)$ in Lemma 2.1 can be represented as*

$$\eta(t) = \int_{-\infty}^{\infty} e^{itx}k(x)\, dZ(x), \tag{2.6}$$

*$Z(x)$ being the random spectral function of $\mathcal{E}(t)$.*

If $f(t)$ is a numerical function such that

$$\int_{-\infty}^{\infty} f(t-s)\, dL(s)$$

exists for every $t$ as an absolutely convergent Riemann-Stieltjes integral, and $X(t) = f(t) + \mathcal{E}(t)$, then we define

$$\int_{-\infty}^{\infty} x(t-s)\, dL(s) = \int_{-\infty}^{\infty} \mathcal{E}(t-s)\, dL(s) + \int_{-\infty}^{\infty} f(t-s)\, dL(s).$$

**3. Convergence theorems.** In this section we shall consider processes of the type

$$Y_n(t) = n\int_{-\infty}^{\infty} X(t-s)K(ns)\, ds, X(t) = f(t) + \mathcal{E}(t) \tag{3.1}$$

and discuss the convergence (in the mean) of $Y_n(t)$ as $n \to \infty$. Similar discussions are classical when $\mathcal{E}(t) \equiv 0$ in the theory of the Fourier integral; for example we have the following fact which we shall state as

LEMMA 3.1:[3] *Suppose that*

(i) $\dfrac{f(s)}{1+|s|} \varepsilon L_1(-\infty, \infty)$,

(ii) $K(s) \varepsilon L_1(-\infty, \infty)$
*and*
(iii) $K(s) = o(|s|^{-1})$ *when* $|s| \to \infty$, *and $K(s)$ is bounded. Then one has*

$$\lim_{n\to\infty} n\int_{-\infty}^{\infty} f(t-s)K(ns)\, ds = f(t)\int_{-\infty}^{\infty} K(s)\, ds. \tag{3.2}$$

---

[3] S. Bochner [1], S. Bochner–K. Chandresekharan [2]. More general theorems are known. See S. Bochner and S. Izumi [3].

Appealing to this lemma we have the following theorem.

THEOREM 3.1: *Let $f(t)$ and $K(u)$ satisfy the conditions of Lemma 3.1. Then we have*

$$\text{l.i.m.}_{n\to\infty}\, n \int_{-\infty}^{\infty} X(t-s)K(ns)\,ds = X(t)\int_{-\infty}^{\infty} K(s)\,ds. \tag{3.3}$$

The proof of this theorem will be omitted since it is very similar to and easier than the one of Theorem 3.2 later.

If we want to estimate the error between both sides of (3.3) for instance as $o(1/n)$, it is necessary to prove a convergence theorem which contains an error term such as the following lemma:

LEMMA 3.2: *Suppose that*

(i) $$\frac{f(s)}{1+|s|^{3/2}} \,\varepsilon\, L_1(-\infty,\infty),$$

(ii) $f(t+u) - f(t) = O(u)$
*for small* $u$,
(iii) $(1+|s|)K(s)\,\varepsilon\, L_1$ *and*
(iv) $K(s)$ *is bounded and* $o(|s|^{-3/2})$ *as* $|s|\to\infty$.
*Then one has*

$$n\int_{-\infty}^{\infty} f(t-s)K(s)\,ds = f(t)\int_{-\infty}^{\infty} K(s)\,ds + o(1/\sqrt{n}). \tag{3.4}$$

PROOF: Put

$$I = n\int_{-\infty}^{\infty} f(t-s)K(ns)\,ds - f(t)\int_{-\infty}^{\infty} nK(ns)\,ds.$$

We want to prove

$$I = o(1/\sqrt{n}). \tag{3.5}$$

We have

$$I = n\int_{-\infty}^{\infty} [f(t-s) - f(t)]K(ns)\,ds$$

$$= n\int_{|s|<\frac{\alpha}{n^{1/2}}} [f(t-s) - f(t)]K(ns)\,ds$$

$$+ n\int_{|s|>\alpha/n^{1/2}} f(t-s)K(ns)\,ds - nf(t)\int_{|s|>\alpha/n^{1/2}} K(ns)\,ds$$

$$= I_1 + I_2 + I_3,$$

say, where $\alpha$ is an arbitrary positive number fixed for the moment.

By (iii), we have

$$
(3.5)\qquad \begin{aligned} I_3 &= O\left(n \cdot \frac{n^{1/2}}{\alpha} \int_{|s| > \alpha/n^{1/2}} | sK(ns) | \, ds\right) \\ &= O\left(\frac{1}{\alpha n^{1/2}} \int_{|u| > \alpha n^{1/2}} | uK(u) | \, du\right) = o\left(\frac{1}{\alpha n^{1/2}}\right) \end{aligned}
$$

as $n \to \infty$.
By (ii), we have

$$
(3.6)\qquad \begin{aligned} I_1 &= O\left(n \int_{|s| < \alpha/n^{1/2}} | sK(ns) | \, ds\right) = O\left(\alpha n^{1/2} \int_{|s| < \frac{\alpha}{n^{1/2}}} | K(ns) | \, ds\right) \\ &= O\left(\frac{\alpha}{n^{1/2}} \int_{-\infty}^{\infty} | K(u) | \, du\right) = O\left(\frac{\alpha}{n^{1/2}}\right). \end{aligned}
$$

Lastly we have, by making use of (iv) and (i)

$$
(3.7)\qquad \begin{aligned} I_2 &= O\left(n \int_{|s| > \alpha/n^{1/2}} | f(t - s) | \, | K(ns) | \, ds\right) \\ &= o\left(\frac{1}{n^{1/2}} \int_{|s| > \alpha/n^{1/2}} | f(t - s) | \frac{ds}{(1 + | s |)^{3/2}}\right) \\ &= o\left(\frac{1}{n^{1/2}} \int_{-\infty}^{\infty} \frac{| f(t - s) |}{(1 + | s |)^{3/2}} \, ds\right) = o\left(\frac{1}{n^{1/2}}\right). \end{aligned}
$$

Combining (3.5), (3.6) and (3.7), we get

$$
\limsup_{n \to \infty} \sqrt{n} I = O(\alpha).
$$

Since $\alpha$ is arbitrary, we must have

$$
\lim_{n \to \infty} \sqrt{n} I = 0,
$$

which proves the lemma.

We shall prove

THEOREM 3.2: *If the conditions* (i), (ii), (iii) and (iv) *of Lemma* 3.2 *are satisfied, and the spectral function $F(x)$ of a continuous stationary process $\mathcal{E}(t)$ satisfies*

$$
(3.8)\qquad \int_{-\infty}^{\infty} | x | \, dF(x) < \infty,
$$

*then*

$$
(3.9)\qquad E\left| n \int_{-\infty}^{\infty} X(t - s)K(ns) \, ds - X(t) \int_{-\infty}^{\infty} K(s) \, ds \right|^2 = o\left(\frac{1}{n}\right),
$$

*where $X(t) = f(t) + \mathcal{E}(t)$.*

PROOF: We have

$$I = n \int_{-\infty}^{\infty} X(t - s)K(ns)\, ds - X(t)n \int_{-\infty}^{\infty} K(ns)\, ds$$

$$= n \int_{-\infty}^{\infty} [X(t - s) - X(t)]K(ns)\, ds$$

$$= n \int_{-\infty}^{\infty} [\mathcal{E}(t - s) - \mathcal{E}(t)]K(ns)\, ds + n \int_{-\infty}^{\infty} [f(t - s) - f(t)]K(ns)\, ds$$

$$= I_1 + I_2,$$

say. Since $EI_1 = 0$, we have

$$E \mid I \mid^2 = E \mid I_1 \mid^2 + \mid I_2 \mid^2.$$

Lemma 3.2 shows $\mid I_2 \mid^2 = o(1/n)$. Hence it is sufficient to show that

$$E \mid I_1 \mid^2 = o(1/n) \tag{3.10}$$

We may now write

$$I_1 = n \int_{-\infty}^{\infty} \mathcal{E}(t - s)K(ns)\, ds - \int_{-\infty}^{\infty} \mathcal{E}(t - s)\, du(s) \cdot \int_{-\infty}^{\infty} K(s)\, ds$$

where $u(s) = 0$ for $s < 0$, $= 1$ for $s > 0$. $u(s)$ has the Fourier-Stieltjes transform identically equal to 1. Hence by Lemma 2.1 (2.3), we have

$$I_1 = \int_{-\infty}^{\infty} \mathcal{E}(t - s)\, d\left(n \int_{-\infty}^{s} K(n\xi)\, d\xi - u(s) \int_{-\infty}^{\infty} K(\xi)\, d\xi\right),$$

$$E\,|I_1|^2 = \int_{-\infty}^{\infty} \left| n \int_{-\infty}^{\infty} (e^{-ixs} - 1)K(ns)\, ds \right|^2 dF(x).$$

Minkowski's inequality shows

$$E\,|I_1|^2 \leqq \left(n \int_{-\infty}^{\infty} |K(ns)|\, ds \left(\int_{-\infty}^{\infty} |e^{-ixs} - 1|^2\, dF(x)\right)^{1/2}\right)^2,$$

which we write as

$$\left(n \int_{-\infty}^{\infty} |K(ns)| \left\{s^2 \int_{-G}^{G} |x|^2\, dF(x) + 2 \int_{|x|>G} |xs|\, dF(x)\right\}^{1/2} ds\right)^2,$$

$G$ being an arbitrary positive number. This does not exceed

$$\left[n \int_{-\infty}^{\infty} |s|\,|K(ns)|\,ds \left(\int_{-G}^{G} |x|^2\,dF(x)\right)^{1/2} + 2^{1/2} n \int_{-\infty}^{\infty} |s|^{1/2}\,|K(ns)|\,ds \left(\int_{|x|>G} |x|\,dF(x)\right)^{1/2}\right]^2$$

$$= \left[\frac{1}{n} \int_{-\infty}^{\infty} |uK(u)|\,du \left(\int_{-G}^{G} |x|^2\,dF(x)\right)^{1/2} + 2^{1/2} \frac{1}{n^{1/2}} \int_{-\infty}^{\infty} |uK(u)|\,du \left(\int_{|x|>G} |x|\,dF(x)\right)^{1/2}\right]^2.$$

Hence we have

$$\limsup_{n\to\infty} nE\,|I_1|^2 = O\left(\int_{|x|>G} |x|\,dF(x)\right)$$

which proves (3.10), since $G$ may be arbitrarily large. Thus the theorem is proved.

If further conditions are imposed on $f(t)$ and $F(x)$, then we can go further and get the asymptotic expression of $\int_{-\infty}^{\infty} X(t-s)\,dK(s)$. We shall leave this until another occasion.

**4. Wiener's formula.** Wiener was concerned with the formula

$$\lim_{a\to\infty} \int_{-\infty}^{\infty} f(t) aK(at)\,dt = \int_{-\infty}^{\infty} K(x)\,dx \cdot \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)\,dt,$$

under suitable conditions on $f(t)$ and $K(t)$. We shall consider the similar formula concerning a stationary process. Let $X(t) = f(t) + \mathcal{E}(t)$ as in the preceding sections. It seems convenient first of all to state a remark.

It is known as the law of large numbers that $(1/2T) \int_{-T}^{T} \mathcal{E}(t)\,dt$ is convergent in mean as $T \to \infty$ and actually

$$\underset{T\to\infty}{\text{l.i.m.}} \frac{1}{2T} \int_{-T}^{T} \mathcal{E}(t) e^{-i\xi t}\,dt = Z(\xi + 0) - Z(\xi - 0),$$

where $\xi$ is any number and $Z(x)$ is the random spectral function. This is also well known (Doob [5]). Hence if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t) e^{-i\xi t}\,dt = M_\xi$$

exists for some $\xi$, then

$$\underset{T\to\infty}{\text{l.i.m.}} \frac{1}{2T} \int_{-T}^{T} X(t) e^{-i\xi t}\,dt = Z(\xi + 0) - Z(\xi - 0) + M_\xi. \tag{4.1}$$

Now we consider

$$\int_{-\infty}^{\infty} \mathcal{E}(t)e^{-i\xi t}aK(at)\,dt. \tag{4.2}$$

Then it is easy to show

THEOREM 4.1: *If $K(t) \varepsilon L_1(-\infty, \infty)$, then*

$$\underset{a\to 0}{\text{l.i.m.}} \int_{-\infty}^{\infty} \mathcal{E}(t)e^{-i\xi t}aK(at)\,dt = [Z(\xi + 0) - Z(\xi - 0)] \int_{-\infty}^{\infty} K(t)\,dt. \tag{4.3}$$

For putting the representation (1.3) into (4.2), assuming $\xi = 0$ without loss of generality, we have

$$\int_{-\infty}^{\infty} \mathcal{E}(t)aK(at)\,dt = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{\frac{ixt}{a}} K(t)\,dt\right) dZ(x),$$

and $\int_{-\infty}^{\infty} e^{ixt/a}k(t)\,dt$ tends to zero boundedly as $a \to 0$ when $x \neq 0$ by the Riemann-Lebesgue lemma and is $\int_{-\infty}^{\infty} K(t)\,dt$ when $x = 0 (a \neq 0)$. Here we used the fact that if $\int_{-\infty}^{\infty} | g_a(x) - g(x)|^2\,dF(x) \to 0$, then

$$\int_{-\infty}^{\infty} g_a(x)\,dZ(x) \to \int_{-\infty}^{\infty} g(x)\,dZ(x).$$

Now a Wiener-type formula of S. Bochner's states [1]: if

(i) $K(x)$ is absolutely continuous in every finite interval,

(4.4) (ii) $| x^2K(x) | < H$, $K(x) \varepsilon L_1(-\infty, \infty)$, $H$ being a constant,

(4.5) (iii) $\frac{1}{2T}\int_{-T}^{T} |f(t)|\,dt \leqq G$, $G$ being a constant, and

(iv) $M = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)\,dt$ exists,

then

$$\lim_{a\to 0} \int_{-\infty}^{\infty} f(t)aK(at)\,dt = M \int_{-\infty}^{\infty} K(t)\,dt. \tag{4.6}$$

This fact and Theorem 4.1 show immediately that

$$\underset{a\to 0}{\text{l.i.m.}} \int_{-\infty}^{\infty} X(t)e^{-i\xi t}aK(at)\,dt = [M_\xi + Z(\xi + 0) - Z(\xi - 0)] \int_{-\infty}^{\infty} K(t)\,dt. \tag{4.7}$$

From (4.7) and (4.1) the following theorem follows immediately

THEOREM 4.2: *If conditions* (i), (ii) *and* (iii) *above are satisfied and*

$$\frac{1}{2T}\int_{-T}^{T} f(t)e^{-i\xi t}\,dt$$

*exists for some $\xi$, then*

$$\underset{a\to 0}{\text{l.i.m.}} \int_{-\infty}^{\infty} X(t)e^{-i\xi t}aK(at)\,dt = \underset{T\to\infty}{\text{l.i.m.}} \frac{1}{2T}\int_{-T}^{T} X(t)e^{i\xi t}\,dt \int_{-\infty}^{\infty} K(t)\,dt. \tag{4.8}$$

Formula (4.8) means that the both sides exist and are equal.

**5. Periodogram.** Let $X(t) = f(t) + \mathcal{E}(t)$ as before. We suppose in this section that the spectral function $F(x)$ of $\mathcal{E}(t)$ is absolutely continuous and we denote the spectral density as $p(x)$:

$$\int_{-\infty}^{x} p(x)\, dx = F(x).$$

It is known and easily proved that

$$\lim_{T\to\infty} E \frac{1}{4\pi T} \left| \int_{-T}^{T} \mathcal{E}(t) e^{ixt}\, dt \right|^2 = p(x), \tag{5.1}$$

provided $p(x)$ is continuous at $x$.

Now we suppose that

$$\lim_{T\to\infty} \frac{1}{4\pi T^{1-\alpha}} \int_{-T}^{T} f(t) e^{-ixt}\, dt = M_{x,\alpha} \tag{5.2}$$

exists for some $x$ and for some $0 \leqq \alpha < 1$.

Then we have

$$E \frac{1}{4\pi T} \left| \int_{-T}^{T} X(t) e^{-ixt}\, dt \right|^2 = \frac{1}{4\pi T} E \left| \int_{-T}^{T} \varepsilon(t) e^{-ixt}\, dt \right|^2 + \frac{1}{4\pi T} \left| \int_{-T}^{T} f(t) e^{-ixt}\, dt \right|^2.$$

Hence we get, letting $T \to \infty$,

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{4\pi T} E \left| \int_{-T}^{T} X(t) e^{-ixt}\, dt \right|^2 \qquad & \\ = p(x), \qquad & \text{if} \quad \alpha > \tfrac{1}{2}, \\ = p(x) + M_{x,\frac{1}{2}}, \qquad & \text{if} \quad \alpha = \tfrac{1}{2}, \\ = \infty, \qquad & \text{if} \quad 0 \leqq \alpha < \tfrac{1}{2} \quad \text{and} \quad M_{x,\alpha} \neq 0. \end{aligned} \tag{5.3}$$

Now we consider the mean convergence of

$$\frac{1}{4\pi T} \left| \int_{-T}^{T} X(t) e^{-ixt}\, dt \right|^2 \tag{5.4}$$

when $T \to \infty$. We shall call (5.4) the periodogram of $X(t)$, mentioning that this is exactly the periodogram of $f(t)$ if $f(t)$ is a trigonometric polynomial and $\mathcal{E}(t) \equiv 0$. Many authors suggest (U. Grenander [6], U. Grenander and M. Rosenblatt [7, 8], Z. A. Lomnicki and S. K. Zaremba [11]) that (5.4) or

$$J(T) = \frac{1}{4\pi T} \left| \int_{-T}^{T} \mathcal{E}(t) e^{-ixt}\, dt \right|^2 \tag{5.5}$$

does not converge to $p(x)$.

We shall discuss in the following sections the behavior of (5.5) and prove that $J(t)$ does not converge in mean to any random variable when $\mathcal{E}(t)$ behaves like a stationary Gaussian process in a certain sense. In the case where $\mathcal{E}(t)$ is stationary Gaussian process U. Grenander and M. Rosenblatt gave extensive discussions (e.g. U. Grenander and M. Rosenblatt [8]).

**6. Theorems on the periodogram.** We shall impose further conditions on $\mathcal{E}(t)$. We suppose hereafter that $\mathcal{E}(t)$ is real, $E \mid \mathcal{E}(t) \mid^4 < \infty$ for every $t$, and

$$E\mathcal{E}(s)\mathcal{E}(s+u)\mathcal{E}(s+v)\mathcal{E}(s+w) = P(u, v, w), \tag{6.1}$$

is a function of $u$, $v$ and $w$ alone and independent of $s$; that is $\mathcal{E}(t)$ is a stationary process of the fourth order. Further let $P(u, v, w)$ be a continuous function of $u$, $v$ and $w$ in the whole range $R_3$.

Put

$$P(u, v, w) = Q(u, v, w) + P_G(u, v, w), \tag{6.2}$$

where

$$P_G(u, v, w) = \rho(u)\rho(v-w) + \rho(v)\rho(w-u) + \rho(w)\rho(u-v), \tag{6.3}$$

$\rho(u)$ being the covariance function (1.4) of $\mathcal{E}(t)$ as before. If $\mathcal{E}(t)$ is a Gaussian process, then $Q(u, v, w) \equiv 0$. Thus $Q(u, v, w)$ will be considered as a measure of non-Gaussianess and was introduced by Magness (T. A. Magness [12], see also E. Parzen [13]). We also assume that $Q(u, v, w)$ is the Fourier transform of a function $q(\xi', \eta', \zeta')$ which is integrable in $R_3$, bounded, continuous and satisfies the Lipschitz condition at a point $(-\xi, -\xi, \xi)$,

$$Q(u, v, w) = (1/2\pi)^{3/2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q(\xi', \eta', \zeta') e^{-i(u\xi'+v\eta'+w\zeta')}\, d\xi'\, d\eta'\, d\zeta' \tag{6.4}$$

Let $\mathcal{E}(t)$ have the spectral density $p(x)$, assumed to be continuous at $x = \xi$ and bounded.

Under these conditions, we shall prove the following theorem.

THEOREM 6.1:

$$J(T) = \frac{1}{4\pi T} \left| \int_{-T}^{T} \mathcal{E}(t) e^{-i\xi t}\, dt \right|^2$$

*satisfies the limit relation*

$$\lim_{T' > T \to \infty} \left\{ E \mid J(T) - J(T') \mid^2 - \left(1 - \frac{T}{T'}\right) 2p^2(\xi) \right\} = 0 \tag{6.5}$$

*if* $\xi \neq 0$, *and*

$$\lim_{T' > T \to \infty} \left\{ E \mid J(T) - J(T') \mid^2 - \left(1 - \frac{T}{T'}\right) \cdot 4p^2(0) \right\} = 0. \tag{6.6}$$

The theorem implies that $E \mid J(T) - J(T') \mid^2$ never converges; in other words $J(T)$ never converges in mean except at a point $\xi$ where $p(\xi) = 0$.

As a theorem for the covariance of $J(T)$ we get under our assumptions above

THEOREM 6.2: *We have*

$$\lim_{T' > T \to \infty} \left\{ \operatorname{Cov}(J(T), J(T')) - \left(1 + \frac{2T}{T'}\right) p^2(\xi) \right\} = 0 \tag{6.7}$$

if $\xi \neq 0$, and

$$\lim_{T' > T \to \infty} \left\{ \mathrm{Cov}\,(J(T), J(T')) - \frac{2T}{T'} p^2(0) \right\} = 0. \tag{6.8}$$

This follows immediately from the fact that

$$\lim_{T' > T \to \infty} \left\{ EJ(T)J(T') - \left(1 + \frac{T}{T'}\right) p^2(\xi) \right\} = 0, \tag{6.9}$$

if $\xi \neq 0$ and

$$\lim_{T' > T \to \infty} \left\{ EJ(T)J(T') - \left(1 + \frac{2T}{T'}\right) p^2(0) \right\} = 0. \tag{6.10}$$

The proofs of above theorems will be done in Section 10.

**7. Lemmas.** It seems convenient to state lemmas in advance which will be used in the courses of proofs of the theorems.

LEMMA 7.1: *Let $p(x) \in L_1(-\infty, \infty)$ and be continuous. Then we have*

$$\lim_{T \to \infty} \frac{1}{\pi} \int_{-\infty}^{\infty} p(x) \frac{\sin T(x + \xi) \sin T(x - \xi)}{T(x + \xi)(x - \xi)}\, dx = p(0), \quad \text{if} \quad \xi = 0 \tag{7.1}$$
$$= 0, \qquad \text{if} \quad \xi \neq 0.$$

The integral when $\xi = 0$ is the Fejér integral and the case $\xi = 0$ is very well known. The case $\xi \neq 0$ was proved by U. Grenander [6]. Some Fourier integral theorems involving the integral (7.1) and having a close connection with estimation theory of the spectral density of a stationary process were discussed by the author recently (T. Kawata [8]).

LEMMA 7.2: *Let $p(x) \in L_1(-\infty, \infty)$ and be continuous. Then we have*

$$\lim_{T' \geq T \to \infty} \left\{ \frac{1}{\pi\sqrt{TT'}} \int_{-\infty}^{\infty} p(x) \frac{\sin T(x + \xi) \sin T'(x + \xi)}{(x + \xi)^2}\, dx - p(\xi) \sqrt{\frac{T}{T'}} \right\} = 0 \tag{7.2}$$

*and*

$$\lim_{T' \geq T \to \infty} \frac{1}{\pi} \frac{1}{\sqrt{TT'}} \int_{-\infty}^{\infty} p(x) \frac{\sin T(x + \xi) \sin T'(x - \xi)}{(x + \xi)(x - \xi)}\, dx = 0, \quad \text{if} \quad \xi \neq 0. \tag{7.3}$$

LEMMA 7.3: *Let $\delta$ be any positive number and let $S(\delta)$ be the domain $|x| < \delta$, $|y| < \delta$, $|z| < \delta$ in $R_3$, the three dimensional Euclidean space. Then*

$$\iiint_{R_3 - S(\delta)} \left| \frac{\sin T(x + y + z)}{x + y + z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z} \right| dx\, dy\, dz = O(T \log^2 T') \tag{7.4}$$

*as $T$ and $T'$ tend to infinity in any way.*

LEMMA 7.4: *If $\varphi(x, y, z)$ is bounded and satisfies*

$$|\varphi(x, y, z)| \leqq C(|x| + |y| + |z|)$$

*for some constant $C$, near the origin, then*

$$\iiint_{-\infty}^{\infty} \left| \varphi(x, y, z) \frac{\sin T(x + y + z)}{x + y + z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z} \right| dx\, dy\, dz \tag{7.5}$$
$$= O(\log^2 T \log T' + T \log^2 T')$$

*as $T$ and $T'$ tend to infinity in such a way that $T' > T$.*

We should like to add a remark. Lemma 7.4 suggests that we should have a convergence theorem like

$$\lim_{T\to\infty} C \iiint_{-\infty}^{\infty} f(x, y, z) \frac{\sin T(x + y + z) \sin Tx \sin Ty \sin Tz}{T(x + y + z)xyz}\, dx\, dy\, dz \tag{7.6}$$
$$= f(0, 0, 0).$$

In fact, under some conditions, (7.5) is true, and a more general theorem was proved by Bochner and the author [4].

We shall prove Lemma 7.2. Since

$$\frac{1}{\pi\sqrt{TT'}} \int_{-\infty}^{\infty} \frac{\sin T(x + \xi) \sin T'(x + \xi)}{(x + \xi)^2}\, dx$$
$$= \frac{1}{\pi\sqrt{TT'}} \int_{-\infty}^{\infty} \frac{\sin T w \sin T'w}{w^2}\, dw = \sqrt{\frac{T}{T'}}, \qquad \text{if} \quad T' \geqq T,$$

we have

$$\frac{1}{\pi\sqrt{TT'}} \int_{-\infty}^{\infty} p(x) \frac{\sin T(x + \xi) \sin T'(x + \xi)}{(x + \xi)^2}\, dx - p(\xi) \sqrt{\frac{T}{T'}}$$
$$= \frac{1}{\pi\sqrt{TT'}} \int_{-\infty}^{\infty} [p(w - \xi) - p(\xi)] \frac{\sin Tw \sin T'w}{w^2}\, dw \tag{7.7}$$
$$= \frac{1}{\pi\sqrt{TT'}} \int_{|w|>\delta} + \frac{1}{\pi\sqrt{TT'}} \int_{|w|<\delta},$$

where $\delta$ is a positive number such that for a given $\epsilon > 0$,

$$|p(w - \xi) - p(\xi)| < \epsilon, \qquad \text{for } |w| < \delta, \tag{7.8}$$

because of the continuity and evenness of $p(x)$. The first term of (7.7) converges to zero as $T, T' \to \infty$, and the second term is less than

$$\frac{\epsilon}{\pi\sqrt{TT'}} \int_{-\infty}^{\infty} \left| \frac{\sin Tw}{w} \frac{\sin T'w}{w} \right| dw$$
$$\leqq \frac{\epsilon}{\pi\sqrt{TT'}} \left( \int_{-\infty}^{\infty} \frac{\sin^2 Tw}{w}\, dw \right)^{\frac{1}{2}} \left( \int_{-\infty}^{\infty} \frac{\sin^2 T'w}{w}\, dw \right)^{\frac{1}{2}} = \epsilon.$$

We shall next prove (7.3). We can easily prove, by the Parseval relation,

$$\frac{1}{\pi\sqrt{TT'}}\int_{-\infty}^{\infty}\frac{\sin T(x+\xi)\sin T'(x-\xi)}{(x+\xi)(x-\xi)}\,dx = \frac{\sin 2T\xi}{\sqrt{TT'}\xi}$$

if $T' \geqq T$, $\xi \gtrless 0$. Hence the left hand side of (7.3) becomes

$$\frac{1}{\pi\sqrt{TT'}}\int_{-\infty}^{\infty}[p(x)-p(\xi)]\,\frac{\sin T(x+\xi)\sin T'(x-\xi)}{(x+\xi)(x-\xi)}\,dx + p(\xi)\,\frac{\sin 2T\xi}{\pi\sqrt{TT'}\xi}$$

Dividing the first integral into three parts as

$$\int_{|x-\xi|<\delta}+\int_{|x+\xi|<\delta}+\int_{|x-\xi|>\delta,|x+\xi|>\delta},$$

$\delta$ being chosen so that $|\delta| < \xi$, and proceeding as the proof of Lemma (7.2), we can prove (7.3).

**8. Proof of Lemma 7.3.** We shall change the notation for simplicity. We write $T_0 = T$, $T_1 = T$, $T_2 = T'$, $T_3 = T'$ and $x_1$, $x_2$, $x_3$ for $x$, $y$, $z$ respectively. Denote $D_0 = [x_i > \delta,\ i = 1, 2, 3]$. The integral in (7.4) is written as

$$I = \iiint_{R3-S(\delta)}\left|\frac{\sin(T_0\sum x_i)}{\sum x_i}\prod_{i=1}^{3}\frac{\sin T_i x_i}{x_i}\right| dv,$$

$dv$ being a volume element in $R_3$, which we divide as

$$(8.1)\qquad I = \iiint_{D_0} + \sum_i \iiint_{D_i} + \sum_{i\neq j}\iiint_{D_{ij}} = I_1 + I_2 + I_3,$$

say, $D_i$ being the domain $D_0 - [\,|x_i| > \delta]$, and $D_{ij}$ being

$$D_0 - [\,|x_i| > \delta,\ |x_j| > \delta].$$

The first integral of the right hand side of (8.1) will be further divided into integrals of four types such as

$$(8.2)\qquad \iiint_{x_1,x_2,x_3>\delta}$$

$$(8.3)\qquad \iiint_{x_i,x_j>\delta;x_k<-\delta}$$

$$(8.4)\qquad \iiint_{x_i>\delta;x_j,x_k<-\delta}$$

$$(8.5)\qquad \iiint_{x_1,x_2,x_3<-\delta},$$

where $i$, $j$, $k$ are distinct. We shall estimate each of the integrals successively. First (8.2) is not greater than

$$\begin{aligned} &\iiint_{x_1,x_2,x_3>\delta} \frac{dv}{x_1\, x_2\, x_3(x_1+x_2+x_3)} \\ &= \int_\delta^\infty \frac{dx_3}{x_3}\int_\delta^\infty \frac{dx_2}{x_2}\int_\delta^\infty \frac{dx_1}{x_1(x_1+x_2+x_3)} = \int_\delta^\infty \frac{dx_3}{x_3}\int_\delta^\infty \frac{\log[(\delta+x_2+x_3)/\delta]}{x_2(x_2+x_3)}\, dx_2 \\ &\leqq C\int_\delta^\infty \frac{dx_3}{x_3}\int_\delta^\infty \frac{\log(x_2+x_3)/\delta}{x_2(x_2+x_3)}\, dx_2 \leqq C\int_\delta^\infty \frac{dx_3}{x_3}\int_\delta^\infty \frac{|\log x_2| + \log}{x_2(x_2+x_3)} \\ &= 2C\int_\delta^\infty \frac{1}{x^2}\, |\log x|\, \log \frac{x+\delta}{\delta}\, dx \end{aligned} \tag{8.6}$$

which is finite. Here $C$ is a constant $C(\delta)$ which may differ on each occurrence. Considering the integral of type (8.3), we shall have, for instance

$$\iiint_{x_1,x_2>\delta,x_3<-\delta}$$

which is

$$\iiint_{x_1,x_2,x_3>\delta} \left| \frac{\sin T_0(x_1+x_2-x_3)}{x_1+x_2-x_3} \prod_{i=1}^{3} \frac{\sin T_i\, x_i}{x_i} \right| dv_x \tag{8.7}$$

The integrand of (8.7) does not exceed $1/(x_1x_2x_3\,|\,x_1 + x_2 - x_3\,|\,)$. If we integrate this over $x_1$, $x_2$, $x_3 > \delta$, $|\,x_1 + x_2 - x_3\,| > \delta/2$, then we see that it is not greater than

$$\int_0^\infty \frac{dx_1}{x_1}\int_\delta^\infty \frac{dx_2}{x_2}\,\frac{1}{(x_1+x_2)}\log\frac{x_1+x_2+\delta/2}{\delta/2} + \int_0^\infty \frac{dx_1}{x_1}\int_\delta^\infty \frac{dx_2}{x_2}\,\frac{1}{(x_1+x_2)}\log\frac{x_1+x_2-\delta/2}{\delta/2}$$

and it can be easily shown, as in the estimation of (8.6), that this is finite.

On the other hand the integrand of (8.7) over the domain $x_1$, $x_2$, $x_3 > \delta$, $|\,x_1 + x_2 - x_3\,| < \delta/2$ does not exceed $T(1/x_1x_2x_3)$, and the integral over the domain can be proved to be $\leqq CT$. Hence it has been shown that $(8.2) = O(T)$.

The integrals of type (8.3) will be $O(T)$, which is also shown easily. Each of the integrals of type (8.4) is just the same as the corresponding integral in (8.3) and the integral (8.4) is the same one as (8.2). Hence we get

$$I_1 = O(T) \tag{8.8}$$

in (8.1).

Next we shall consider $I_2$ in the right hand side of (8.1). For instance

$$\begin{aligned} \iiint_{D_1} &= \iiint_{|x_1|\leqq\delta,|x_2|,|x_3|>\delta} \left| \frac{\sin T_0 \sum x_i}{\sum x_i} \prod_{i=1}^{3} \frac{\sin T_i\, x_i}{x_i} \right| dv_x \\ &= \iiint_{D_{11}} + \iiint_{D_{12}} = I_{21} + I_{22}, \end{aligned} \tag{8.9}$$

say, where $D_{11}$ is the domain $[\,|x_1| \leqq \delta, |x_2| > \delta, |x_3| > \delta, |x_1 + x_2 + x_3| > \delta]$ and $D_{12} = [\,|x_1| \leqq \delta, |x_2| > \delta, |x_3| > \delta, |x_1 + x_2 + x_3| < \delta]$.

If we write

$$I_{21} = \iiint_{D_{11}-[|x_1|<\delta/2]} + \iiint_{D_{11}[\delta/2<x_1<\delta]}, \tag{8.10}$$

then the latter is found to be bounded as in the arguments in the first step of the estimation of (8.7). The first integral does not exceed

$$\int_{|x_1|<\delta/2} \left|\frac{\sin Tx_1}{x_1}\right| dx_1 \int_{|x_2|>\delta} \frac{dx_2}{|x_2|} \int_{|x_3|>\delta, |x_1+x_2+x_3|>\delta} \frac{dx_3}{|x_3(x_1 + x_2 + x_3)|}$$

$$\leqq 2 \int_{|x_1|<\delta/2} \left|\frac{\sin Tx_1}{x_1}\right| dx_1 \int_{|x_2|>\delta} \frac{dx_2}{|x_2|} \int_{|x_3|>\delta, |x_2+x_3|>\delta/2} \frac{dx_3}{|x_3(x_2 + x_3)|}$$

since $|x_1 + x_2 + x_3| > \delta$, $|x_1| < \delta/2$ implies $|x_2 + x_3| > \delta/2$. The last integral is not greater than

$$2 \int_{|x_2|>\delta} \frac{dx_2}{|x_2|} \int_{|x_3|>\delta, |x_2+x_3|>\delta/2} \frac{dx_3}{|x_3(x_2 + x_3)|} \cdot \int_{|x_1|<\delta/2} \left|\frac{\sin Tx_1}{x_1}\right| dx_1$$

in which the first factor is $O(1)$ as in the evaluation of (8.7) and the second integral is $O(\log T)$ as is known since it is the Lebesgue constant. Thus

$$I_{21} = O(\log T). \tag{8.11}$$

Next $I_{22}$ will be computed, being written as

$$I_{22} = \iiint_{D_{11}\cdot[|x_1+x_2+x_3|<1/T]} + \iiint_{D_{11}\cdot[\delta>|x_1+x_2+x_3|>1/T]} \tag{8.12}$$

We consider the integral over the domain $D_{111}$: $|x_1| < \delta/2$, $|x_2| > \delta$, $|x_3| > \delta$, $|x_1 + x_2 + x_3| < 1/T$,

$$\iiint_{D_{111}} \tag{8.13}$$

in place of the first integral in the right side of (8.12). The remaining part of the integral can be estimated in the same way as was done in $I_1$, to be $O(T)$.

Thus we have

$$\iiint_{D_{111}} \leqq T \iiint_{D_{111}} \left|\prod_{i=1}^{3} \frac{\sin T_i x_i}{x_i}\right| dv_x$$

$$\leqq T \int_{|x_1|<\delta/2} \left|\frac{\sin Tx_1}{x_1}\right| dx_1 \iint_{\substack{|x_2|,|x_3|>\delta \\ |x_1+x_2+x_3|<1/T}} \left|\frac{\sin T'x_2}{x_2} \frac{\sin T'x_3}{x_3}\right| dx_2\, dx_3$$

$$= 2T \int_{|x_1|<\delta/2} \left|\frac{\sin Tx_1}{x_1}\right| dx_1 \iint_{\substack{x_2>\delta, x_3<-\delta \\ |x_1+x_2+x_3|<1/T}} \left|\frac{\sin T'x_2}{x_2} \frac{\sin T'x_3}{x_3}\right| dx_2\, dx_3$$

since $x_2 > \delta$, $x_3 > \delta$, $|x_1 + x_2 + x_3| < 1/T$ is impossible for large $T$. The last integral does not exceed

$$2T \int_{|x_1|<\delta/2} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 \int_\delta^\infty \frac{dx_2}{x_2} \int_{-x_2-x_1-1/T} \frac{dx_3}{x_3}$$
$$= 2T \int_{|x_1|<\delta/2} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 \int_\delta^\infty \frac{1}{x_1} \log\left(1 + \frac{2}{T}\frac{1}{x_2 + x_1 - 1/T}\right) dx_2$$
$$\leqq CT \int_{|x_1|<\delta/2} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 \frac{1}{T} \int_\delta^\infty \frac{dx_2}{x_2(x_2 - \delta - 1/T)}$$
$$\leqq C \int_{|x_1|<\delta/2} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 = O(\log T).$$

Hence we get that the first integral of the right hand side of $I_{22}$ is $O(T)$.

We next consider the second integral of $I_{22}$ in (8.12), which is not greater than

$$(8.14) \qquad 2 \int_{|x_1|<\delta/2} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 \iint_{\substack{x_2>\delta, x_3<-\delta \\ 1/T<|x_1+x_2+x_3|<\delta}} \left| \frac{1}{(x_1 + x_2 + x_3)x_2 x_3} \right| dx_2\, dx_3 .$$

The inner integral, by a change of a variable, becomes

$$\iint_{\substack{x_2,x_3>\delta \\ 1/T<|x_1+x_2-x_3|<\delta}} \frac{dx_2\, dx_3}{|x_1 + x_2 - x_3|\, x_2 x_3}$$

which is not greater than the sum

$$\int_\delta^\infty \frac{dx_2}{x_2} \int_{x_1+x_2+1/T}^\infty \frac{dx_3}{(x_3 - x_1 - x_2)x_3} + \int_\delta^\infty \frac{dx_2}{x_2} \int_{x_1+x_2-1/T>x_3} \frac{dx_3}{(x_1 + x_2 - x_3)x_3}$$
$$\leqq C \int_{x_1+x_2>\delta, x_2>\delta} \frac{1}{x_2(x_1 + x_2)} \log T(x_1 + x_2)\, dx_2 .$$

This is easily proved to be $O(\log T)$ and hence (8.12) is $O(\log^2 T)$, Lebesgue's constant being involved. Hence we get

$$I_{22} = O(T) + O(\log^2 T) = O(T).$$

Inserting this result and (8.10) into (8.9), we have shown $\iiint_{D_1} = O(T)$. Similar arguments show that

$$\iiint_{D_2} = O(T) + O(\log T \log T')$$

and

$$\iiint_{D_3} = O(T) + O(\log T \log T').$$

The domain $D_2$ and $D_3$ are defined analogously to $D_1$ and the above estimates are easily verified. Combining these results, we get

$$I_2 = O(T \log T'). \tag{8.14}$$

Lastly we shall consider $I_3$. We shall treat, for instance, the integral over $D_{12}$, that is,

$$\begin{aligned} \iiint_{D_{12}} &= \iiint_{|x_1|<\delta, |x_2|<\delta, |x_3|>\delta} \left| \frac{\sin T(\sum x_i)}{\sum x_i} \prod_{i=1}^{3} \frac{\sin T_i x_i}{x_i} \right| dx_1\, dx_2\, dx_3 \\ &= \iiint_{|x_1|<\delta, |x_2|<\delta, |x_3|>3\delta} + \iiint_{|x_1|<\delta, |x_2|<\delta, \delta<|x_3|<3\delta} . \end{aligned} \tag{8.15}$$

Replacing $| x_1 + x_2 + x_3 |$ by $\frac{1}{3} | x_3 |$, because of

$$| x_1 + x_2 + x_3 | > | x_3 | - | x_1 | - | x_2 | > \tfrac{1}{3} | x_3 | ,$$

we see that the first integral does not exceed

$$2 \int_{|x_1|<\delta} \left| \frac{\sin x_1 T}{x_1} \right| dx_1 \int_{|x_2|<\delta} \left| \frac{\sin x_2 T'}{x_2} \right| dx_2 \int_{|x_3|>3\delta} dx_3$$

This is clearly

$$O(\log T \log T'). \tag{8.16}$$

The latter integral of (8.15) is not greater than

$$\begin{aligned} &\frac{1}{\delta} \int_{|x_1|<\delta} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 \int_{|x_2|<\delta} \left| \frac{\sin T'x_2}{x_2} \right| dx_2 \int_{\delta<|x_3|<3\delta} \left| \frac{\sin T(x_1 + x_2 + x_3)}{x_1 + x_2 + x_3} \right| dx_3 \\ &\qquad \leqq \frac{1}{\delta} \int_{|x_1|<\delta} \left| \frac{\sin Tx_1}{x_1} \right| dx_1 \int_{|x_2|<\delta} \left| \frac{\sin T'x_2}{x_2} \right| dx_2 \int_{|u|<5\delta} \left| \frac{\sin Tu}{u} \right| du \\ &\qquad = O\,(\log^2 T \cdot \log T'). \end{aligned}$$

We have thus reached $\iiint_{D_{12}} = O(\log^2 T \log T')$. The other integrals in $I_3$ may be shown by similar arguments to be $O(\log^2 T \log T')$ or $O(\log^2 T' \cdot \log T)$. Hence, combining these results, we get

$$I_3 = O(\log^2 T \log T' + \log^2 T' \cdot \log T). \tag{8.17}$$

By (8.8), (8.14) and (8.17), we finally get $I = O(T \log^2 T')$.

**9. Proof of Lemma 7.4.** Let

$$| \varphi(x, y, z) | \leqq C( | x | + | y | + | z | ) \tag{9.1}$$

in $S(\delta): | x |, | y |, | z | < \delta$. Let $M$ be the upper bound of $\varphi(x, y, z)$. We partition the integral as

$$\iiint_{-\infty}^{\infty} \left| \varphi(x, y, z) \frac{\sin T(x+y+z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z} \right| dx\, dy\, dz$$

$$= \iiint_{S(\delta)} + \iiint_{E_3 - S(\delta)} = J_1 + J_2,$$

say. Then

$$\text{(9.2)} \quad \begin{aligned} |J_2| &\leqq M \iiint_{E_3 - S(\delta)} \left| \frac{\sin T(x+y+z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z} \right| dx\, dy\, dz \\ &= O(T \log^2 T') \end{aligned}$$

by Lemma 7.3.

Next, inserting the relation (9.1), we have

$$\text{(9.3)} \quad \begin{aligned} |J_1| \leqq C \iiint_{S(\delta)} (|x| + |y| + |z|) &\left| \frac{\sin T(x+y+z)}{x+y+z} \right| \\ &\cdot \left| \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z} \right| dx\, dy\, dz \Bigg). \end{aligned}$$

We consider, for instance, the following part of this integral:

$$\begin{aligned} \iiint_{S(\delta)} |y| &\left| \frac{\sin T(x+y+z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z} \right| dx\, dy\, dz \\ &\leqq \iiint_{S(\delta)} \left| \frac{\sin T(x+y+z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'z}{z} \right| dx\, dy\, dz \\ &= \int_{-\delta}^{\delta} \left| \frac{\sin Tx}{x} \right| dx \int_{-\delta}^{\delta} \left| \frac{\sin T'z}{z} \right| dz \int_{-\delta}^{\delta} \left| \frac{\sin T(x+y+z)}{x+y+z} \right| dy \\ &\leqq \int_{-\delta}^{\delta} \left| \frac{\sin Tx}{x} \right| dx \int_{-\delta}^{\delta} \left| \frac{\sin T'z}{z} \right| dz \int_{-3\delta}^{3\delta} \left| \frac{\sin Tu}{u} \right| du \\ &= O\,(\log^2 T \log T'). \end{aligned}$$

Other parts in (9.3) may be estimated by similar arguments to be

$$O(\log T \log^2 T') \qquad \text{or} \qquad O(\log^2 T \log T').$$

Keeping (9.2) in mind, we get the proof of Lemma 7.4.

**10. Proofs of theorems.** We are now in a position to prove the theorems in Section 6. Throughout this section we of course assume all conditions stated in these.

First of all we shall evaluate $EJ^2(T)$.

$$EJ^2(T) = \frac{1}{16\pi^2} \frac{1}{T^2} E \iiiint_{-T}^{T} \mathcal{E}(s)\mathcal{E}(t)\mathcal{E}(u)\mathcal{E}(v) \cdot e^{-i(s-t)\xi + i(u-v)\xi}\, ds\, dt\, du\, dv.$$

Inserting (6.2) here, we have

$$EJ^2(T) = \frac{1}{16\pi^2}\frac{1}{T^2}\int\int\int\int_{-T}^{T} Q(t-s, u-s, v-s)$$

$$\cdot e^{-i\{(s-t)-(u-v)\}\xi}\, ds\, dt\, du\, dv \tag{10.1}$$

$$+ \frac{1}{16\pi^2}\frac{1}{T^2}\int\int\int\int_{-T}^{T} P_G(t-s, u-s, v-s) e^{-i\{(s-t)-(u-v)\}\xi}\, ds\, dt\, du\, dv$$

say. Because of (6.4), we have

$$J_1 = \frac{1}{16\pi^2 T^2}\left(\frac{1}{2\pi}\right)^{3/2}\int\int\int_{-\infty}^{\infty} q(x, y, z)\, dx\, dy\, dz$$

$$\cdot \int\int\int\int_{-T}^{T} e^{-is(x+y+z+\xi)} \cdot e^{it(x+\xi)+iu(y+\xi)+iv(z-\xi)}\, ds\, dt\, du\, dv$$

$$= \left(\frac{1}{2\pi}\right)^{3/2}\frac{1}{\pi^2 T^2}\int\int\int_{-\infty}^{\infty} q(x, y, z)\frac{\sin T(x+y+z+\xi)}{x+y+z+\xi}$$

$$\cdot \frac{\sin T(x+\xi)}{x+\xi}\cdot\frac{\sin T(y+\xi)}{y+\xi}\cdot\frac{\sin T(z-\xi)}{z-\xi}\, dx\, dy\, dz \tag{10.2}$$

$$= \left(\frac{1}{2\pi}\right)^{3/2}\frac{1}{\pi^2 T^2}\int\int\int_{-\infty}^{\infty} q(x-\xi, y-\xi, z+\xi)\frac{\sin T(x+y+z)}{x+y+z}$$

$$\cdot \frac{\sin Tx}{x}\frac{\sin Ty}{y}\frac{\sin Tz}{z}\, dx\, dy\, dz.$$

Now the integral (10.2) with $q(x, y, z) \equiv 1$ is easily shown to be $\pi^3 T$ by the repeated applications of Parseval's relation. Hence we have

$$J_1 = \left(\frac{1}{2\pi}\right)^{3/2}\frac{1}{\pi^2 T^2}\int\int\int_{-\infty}^{\infty}\{q(x-\xi, y-\xi, z+\xi) - q(-\xi, -\xi, \xi)\}$$

$$\cdot \frac{\sin T(x+y+z)}{x+y+z}\frac{\sin Tx}{x}\frac{\sin Ty}{y}\frac{\sin Tz}{z}\, dx\, dy\, dz \tag{10.3}$$

$$+ g(-\xi, -\xi, \xi)\left(\frac{1}{2\pi}\right)^{3/2}\frac{\pi}{T}$$

Since $q(x, y, z)$ satisfies the Lipschtz condition at the point $(-\xi, -\xi, \xi)$, we have, by Lemma 7.4 with $T = T'$,

$$J_1 = O\left(\frac{1}{T}\log^2 T\right) + O\left(\frac{1}{T}\right) \tag{10.4}$$

$$= o(1).$$

Next

$$
(10.5)\qquad
\begin{aligned}
J_2 = \frac{1}{16\pi^2}\frac{1}{T^2}\iiint\int_{-T}^{T} &\cdot\{\rho(t-s)\rho(u-v)+\rho(u-s)\rho(v-t)+\rho(v-s)\rho(t-u)\}\\
&\cdot e^{-i\{(s-t)-(u-v)\}}\,ds\,dt\,du\,dv.
\end{aligned}
$$

Here we shall have, for instance,

$$
(10.6)\qquad \frac{1}{16\pi^2}\frac{1}{T^2}\iiint\int_{-T}^{T}\rho(t-s)\rho(u-v)e^{-i\{(s-t)-(u-v)\}\xi}\,ds\,dt\,du\,dv \to p^2(\xi),\qquad \text{as}\quad T\to\infty,
$$

because this will become, inserting $\rho(u) = \int_{-\infty}^{\infty} p(x)e^{iux}\,dx$, and making a change of orders of integration,

$$
\begin{aligned}
\frac{1}{16\pi^2}\frac{1}{T^2}\int_{-\infty}^{\infty}p(x)\,dx\int_{-\infty}^{\infty}p(y)\,dy\iiint\int_{-T}^{T}e^{i(t-s)(x+\xi)+i(u-v)(y+\xi)}\,dt\,ds\,du\,dv\\
=\int_{-\infty}^{\infty}p(x)\frac{\sin^2 T(x+\xi)}{\pi T(x+\xi)^2}\,dx\cdot\int_{-\infty}^{\infty}p(y)\frac{\sin^2 T(y+\xi)}{\pi T(y+\xi)^2}\,dy
\end{aligned}
$$

which tends obviously to $p^2(-\xi) = p^2(\xi)$ as $T\to\infty$ by the well known property of Fejer's integral.

As for the second part of the integral of the right hand side of (10.5), we obtain

$$
(10.7)\qquad \frac{1}{16\pi^2T^2}\iiint\int_{-T}^{T}\rho(u-s)\rho(v-t)e^{-i\{(s-t)-(u-v)\}}\,ds\,dt\,du\,dv\to p^2(\xi),\qquad \text{as}\quad T\to\infty.
$$

However we find a difference in considering the remaining part. In fact the integral

$$
\frac{1}{16\pi^2T^2}\iiint\int_{-\infty}^{\infty}\rho(v-s)\rho(t-u)e^{-i\{(s-t)-(u-v)\}}\,ds\,dt\,du\,dv
$$

becomes, after similar treatments,

$$
\left[\int_{-\infty}^{\infty}p(x)\frac{\sin T(x+\xi)\sin T(x-\xi)}{\pi T(x+\xi)(x-\xi)}\,dx\right]^2,
$$

which converges to $p^2(0)$ if $\xi = 0$, and 0 if $\xi\neq 0$, by Lemma 7.1. Hence combining this result with (10.6), (10.7), we get

$$
\begin{aligned}
\lim_{T\to\infty} J_2 &= 3\,p^2(0), && \text{if } \xi = 0,\\
&= 2\,p^2(\xi), && \text{if } \xi\neq 0.
\end{aligned}
$$

We finally get

$$\lim_{T\to\infty} EJ^2(T) = 3\, p^2(0), \qquad \text{if } \xi = 0, \tag{10.8}$$

$$= 2\, p^2(\xi), \qquad \text{if } \xi \neq 0.$$

Moreover we shall find out here the limit value of $EJ(T)J(T')$. The same method as above leads us to

$$\frac{1}{16\pi^2 TT'} E \int\int_{-T}^{T} \mathcal{E}(s)\mathcal{E}(t)e^{-i(s-t)\xi}\, ds\, dt \int\int_{-T}^{T} \mathcal{E}(u)\mathcal{E}(v)e^{i(u-v)\xi}\, du\, dv$$

$$= \frac{1}{16\pi^2 TT'} \int\int_{-T}^{T} ds\, dt \int\int_{-T'}^{T'} du\, dv \cdot Q(t-s, u-v, v-s)e^{-i[(s-t)-(u-v)]\xi}$$

$$+ \frac{1}{16\pi^2 TT'} \int\int_{-T}^{T} ds\, dt \int\int_{-T'}^{T'} du\, dv P_Q(t-s, u-s, v-s)e^{-i[(s-t)-(u-v)]\xi} \tag{10.9}$$

$$= K_1(T, T') + K_2(T, T')$$

say. By the same way we got (10.2), we shall have

$$K_1(T, T') = \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{\pi^2} \frac{1}{TT'} \int\int\int_{-\infty}^{\infty} q(x-\xi, y-\xi, z+\xi)$$

$$\cdot \frac{\sin T(x+y+z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z}\, dx\, dy\, dz.$$

Now the Parseval relation proves $K_1(T, T')$ with $q(x, y, z) \equiv 1$, to be

$$\int\int\int_{-\infty}^{\infty} \frac{\sin T(x, y, z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z}\, dx\, dy\, dz = \pi^3 T, \tag{10.10}$$

if $T' > T$. Hence we have

$$K_1(T, T') = \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{\pi^2} \frac{1}{TT'} \int\int\int_{-\infty}^{\infty} \{q(x-\xi, y-\xi, z+\xi) - q(-\xi, -\xi, \xi)\}$$

$$\cdot \frac{\sin T(x+y+z)}{x+y+z} \frac{\sin Tx}{x} \frac{\sin T'y}{y} \frac{\sin T'z}{z}\, dx\, dy\, dz$$

$$+ \left(\frac{1}{2\pi}\right)^{3/2} \frac{\pi}{T'} q(-\xi, -\xi, \xi)$$

which is, by Lemma 2,

$$O\left(\frac{1}{TT'}(\log^2 T \log T' + T \log^2 T') + \frac{1}{T'}\right).$$

Hence

$$K_1(T, T') = o(1), \qquad \text{as } T' \geqq T \to \infty. \tag{10.11}$$

Next $K_2(T, T')$ will become after some arguments

$$\frac{1}{16\pi^2}\frac{1}{TT'}\iint_{-T}^{T} ds\,dt \iint_{-T'}^{T'} du\,dv\rho(t-s)\rho(u-v)e^{-i[(s-t)-(u-v)]\xi}$$

$$+\iint_{-T}^{T} ds\,dt \iint_{-T'}^{T'} du\,dv\rho(u-s)\rho(v-t)e^{-i[(s-t)-(u-v)]\xi}$$

$$+\iint_{-T}^{T} ds\,dt \iint_{-T'}^{T'} du\,dv\rho(v-s)\rho(t-u)e^{-i[(s-t)-(u-v)]\xi}$$

$$= K_{21} + K_{21} + K_{23},$$

say. It will be seen that the asymptotic behaviors are much different among $K_{21}$, $K_{22}$ and $K_{23}$.

In fact $K_{21}$ becomes, by an argument like that used in considering (10,6),

$$\int_{-\infty}^{\infty}\frac{\sin^2 T(x+\xi)}{\pi T(x+\xi)^2}\,p(x)\,dx\int_{-\infty}^{\infty}\frac{\sin^2 T'(y-\xi)}{\pi T'(y-\xi)^2}\,p(y)\,dy,$$

which tends to $p(\xi)p(-\xi) = p^2(\xi)$ as $T, T' \to \infty$. Thus

$$\lim_{T' \geq T \to \infty} K_{21}(T, T') = p^2(\xi). \tag{10.12}$$

Next we see that $K_{22}(T, T')$ will become

$$\frac{1}{\pi^2}\frac{1}{TT'}\int_{-\infty}^{\infty}\frac{\sin T(x+\xi)\sin T'(y+\xi)}{(x+\xi)^2}\,p(x)\,dx$$

$$\cdot\int_{-\infty}^{\infty}\frac{\sin T(y-\xi)\sin T'(y-\xi)}{(y-\xi)^2}\,p(y)\,dy.$$

Then Lemma 7.2, (7.2) shows

$$\lim_{T' \geq T \to \infty}\left\{K_{22}(T, T') - p^2(\xi)\frac{T}{T'}\right\} = 0. \tag{10.13}$$

Finally $K_{23}$ becomes, after some arguments,

$$\frac{1}{\pi^2}\frac{1}{TT'}\left[\int_{-\infty}^{\infty}\frac{\sin(x+\xi)T\sin T'(x-\xi)}{(x+\xi)(x-\xi)}\,p(x)\,dx\right]^2$$

which converges to 0 if $\xi \neq 0$, by Lemma 7.2, (7.3). On the other hand if $\xi = 0$. then it reduces to $K_{22}$ and

$$\lim_{T' \geq T \to \infty}\left\{K_{22}(T, T') - p^2(0)\frac{T}{T'}\right\} = 0.$$

Combining (10.12), (10.13) and the last result, we have

$$(10.14)\qquad \lim_{T' \geq T \to \infty} \left\{ K_2(T, T') - \left(1 + \frac{T}{T'}\right) p^2(\xi) \right\} = 0, \qquad \text{if } \xi \neq 0,$$

$$(10.15)\qquad \lim_{T' \geq T \to \infty} \left\{ K_2(T, T^1) - \left(1 + \frac{2T}{T'}\right) p^2(0) \right\} = 0.$$

Hence putting (10.11), (10.14) and (10.15) into (10.9) we get (6.9) and (6.10).

After these preparations, the proofs of the theorems are now very easy. For

$$\begin{aligned} E\{J(T) - J(T')\}^2 - \left(1 - \frac{T}{T'}\right) 2p^2(\xi) \\ &= (EJ^2(T) - 2p^2(\xi)) + (EJ^2(T') - 2p^2(\xi)) \\ &\quad - 2\left(EJ(T)J(T') - \left(1 - \frac{T}{T'}\right) p^2(\xi)\right), \end{aligned}$$

which tends to zero by virtue of (10.8) and (6.9). Formula (6.6) is also proved using (10.8) and (6.10).

The proof of Theorem 6.2 is also immediate.

## REFERENCES

[1] S. Bochner, *Vorlesungen über Fourierintegrale*, Teubner, Leipzig, 1932.

[2] S. Bochner and K. Chandrasekharan, *Fourier Transforms*, Princeton University Press, 1949.

[3] S. Bochner and S. Izumi, "Some general convergence theorems," *Tohoku Math. J.* Vol. 42 (1936), pp. 191–194.

[4] S. Bochner and T. Kawata, "A special integral transformation in Euclidean space," *Ann. Mathematics*, Vol. 68 (1958), pp. 150–158.

[5] J. L. Doob, *Stochastic Processes*, John Wiley, New York, 1952.

[6] U. Grenander, "On empirical spectral analysis of stochastic processes," *Ark. Mat.* Vol. 1 (1951), pp. 503–531.

[7] U. Grenander and M. Rosenblatt, "Statistical spectral analysis of time series arising from stationary stochastic processes," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 537–558.

[8] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley, New York, 1957.

[9] T. Kawata, "Some Fourier integral theorems," to be published in *Kodai Math. Sem. Rep.*

[10] M. Loève, *Fonctions Aléatoires du Second Ordre*, Supplement to P. Levy, *Processus Stochastiques et Mouvement Brownien*, Gauthier, Paris, 1948.

[11] Z. A. Lomnicki and S. K. Zaremba, "On the estimation of autocorrelation," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 140–158.

[12] T. A. Magness, "Spectral response of a quadratic device to non-Gaussian noise," *J. Appl. Physics*, Vol. 25 (1954), pp. 1357–1365.

[13] E. Parzen, "On consistent estimates of the spectrum of a stationary time series," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 329–348.

# LARGE EXCURSIONS OF GAUSSIAN PROCESSES

By Mark Kac and David Slepian

*Cornell University and Bell Telephone Laboratories, Inc.*

**1. Introduction.** It is known that the problem of determining the distribution of spacings between consecutive $a$-values of an ergodic Gaussian process, $x(t)$, ($Ex(t) = 0$, $Ex^2(t) = 1$) is very difficult. Recently Palmer [1] and Rice [2] treated some limiting cases of this problem. In one limit they determine, for $a \to \infty$, the conditional probability

$$\Pr\{x(\tau) > a, \quad 0 \leqq \tau \leqq t\theta(a) \mid x(0) = a, \quad x'(0) > 0\} \tag{1.1}$$

where $\theta(a)$ is the average length of the times spent by $x(t)$ above the level $a$. Apart from some differences concerning the meaning of the conditional probability (1.1) both authors use the following heuristic device.

Since for large $a$, $\theta(a)$ is small, they write

$$x(\tau) = a + x'(0)\,\tau + \frac{x''(0)}{2}\,\tau^2 \tag{1.2}$$

and take for the time of the first downward crossing of the $a$-level

$$\tau = -2\,\frac{x'(0)}{x''(0)}. \tag{1.3}$$

It would thus seem that this procedure is limited to processes for which $x''$ exists. This would exclude, for example, the displacement of a harmonic oscillator in Brownian motion. It is precisely this point that led us to undertake the present investigation.

We have found an alternative derivation of the Palmer-Rice results which does not depend on the approximation (1.2) and hence is applicable to all cases of physical interest. We have also attempted to elucidate the ambiguity of (1.1) (see §2) and we have in §3 shown in what sense the sample functions $x(\tau)$ are approximated by parabolas as suggested in (1.2).

**2. Conditional probability densities.** It is well known that conditional probabilities and conditional probability densities must frequently be treated with some care. Since the material to follow contains some excellent examples of the subtle nature of these quantities, a few words on the subject are in order here.

Let $x(t)$ be a continuous ergodic Gaussian process possessing a derivative almost everywhere. Consider the "conditional probability density for the slope $\xi = x'(0)$ given that $x(0) = a$." From the ensemble point of view, the phrase in quotation marks has no meaning, since the set of sample functions satisfying the condition $x(0) = a$ is of probability zero. Yet, given a sample function of

Received July 14, 1958; revised December 29, 1958.

the process, one can imagine observing the slope of $x(t)$ at each value of $t$ for which $x(t) = a$ and one can thus obtain an "empirical or time derived probability density for $x'(t)$ given that $x(t) = a$." This probability density will be the same for almost all sample functions. How do we reconcile these two points of view?

From the ensemble point of view, one can, of course, give meaning to the "probability density for $\xi = x'(0)$ given that $x(0) = a$" by means of limiting procedures. The condition $x(0) = a$ is replaced by some condition, $A$, of positive probability depending on parameters. The condition is chosen so that as the parameters assume limiting values, $A$ becomes the condition $x(0) = a$. The conditional density of $\xi$ given $A$, $p(\xi \mid A)$, is computed; the limit of this quantity as the parameters assume their limiting values can then be taken as a definition of $p(\xi \mid x(0) = a)$, the "density for $\xi$ given that $x(0) = a$."

Unfortunately this limit depends in general on the manner in which $A$ approaches the condition $x(0) = a$. We illustrate with a few examples.

(i) Let $A$ be $a \leqq x(0) \leqq a + \delta$. Then

$$p(\xi|x(0) = a)_{\text{v.w.}} = \lim_{\delta\to 0} \frac{\int_a^{a+\delta} p(\xi, x)\, dx}{\int_a^{a+\delta} p(x)\, dx} = \frac{e^{-\frac{\xi^2}{2\alpha}}}{\sqrt{2\pi\alpha}}$$

where the subscript v.w. stands for "vertical window," $p(\xi, x)$ is the joint density for $\xi = x'(0)$ and $x(0)$, and $p(x)$ is the density for $x(0)$. We have made use of the independence of $x(0)$ and $\xi$ and have assumed that $Ex(t) = 0$ and $E\xi^2 = \alpha$. This vertical window definition of the conditional density of $\xi$ given that $x(0) = a$ thus reduces to the conventional one $p(\xi|x(0) = a) = [p(\xi, x)/p(x)]_{x=a}$.

(ii) Let $A$ be the "horizontal window condition" $x(t) = a$ for some $t$ such that $0 \leqq t \leqq \delta$. Then if $\xi \geqq 0$,

$$p(\xi \mid x(0) = a)_{\text{h.w.}}$$

$$= \lim_{\delta\to 0} \frac{\int_{a-\xi\delta}^{a} p(\xi, x)\, dx}{\int_0^{\infty} d\xi' \int_{a-\xi'\delta}^{a} dx\, p(\xi', x) + \int_{-\infty}^{0} d\xi' \int_a^{a-\xi'\delta} dx\, p(\xi', x)} = \frac{\xi}{2\alpha}\, e^{-\frac{\xi^2}{2\alpha}} \tag{2.1}$$

since the condition $A$ can be satisfied (to first order in small quantities) for a given value of slope, say $\xi' > 0$, only if $a - \xi'\delta \leqq x(0) \leqq a$. A similar calculation for $\xi < 0$, gives the final result

$$p(\xi \mid x(0) = a)_{\text{h.w.}} = \frac{|\xi|}{2\alpha}\, e^{-\frac{\xi^2}{2\alpha}}.$$

(iii) More generally, let $A$ be the condition that $x(t)$ pass through a line segment of length $\delta$ and slope $m$ having one end-point at $x = a$, $t = 0$. Then one

finds by straightforward computation

$$p(\xi \mid x(0) = a)_m = \frac{\mid \xi - m \mid e^{-\frac{\xi^2}{2\alpha}}/\sqrt{2\pi\alpha}}{\sqrt{\frac{2\alpha}{\pi}}\, e^{-\frac{m^2}{2\alpha}} + m \int_{-m}^{m} \frac{e^{-\frac{x^2}{2\alpha}}}{\sqrt{2\pi\alpha}}\, dx}.$$

(iv) If $A$ is the condition that $x(t)$ pass through a circle of radius $\delta$ with center at the point $x = a$, $t = 0$, then

$$p(\xi \mid x(0) = a)_0 = \frac{\sqrt{1 + \xi^2}\, e^{-\frac{\xi^2}{2\alpha}}}{\int_{-\infty}^{\infty} \sqrt{1 + x^2}\, e^{-\frac{x^2}{2\alpha}}\, dx}.$$

Which, if any, of these several versions of $p(\xi|x(0) = a)$ is equal to the empirical density obtained from a single sample function? The question can be answered readily in the following heuristic manner. Let $\nu$ be the expected number of zeros per unit time of $x(t) - a$. Let $S_b(y) = 1$ if $y \leqq b$ and be zero otherwise. The empirical cumulative distribution for $\xi$ can be written,

$$\begin{aligned}
\Pr (\xi \leqq b \mid x(0) = a)_{\text{emp}} &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \frac{1}{\nu}\, \delta[x(t) - a] \mid x'(t) \mid S_b[x'(t)]\, dt \\
&= E \frac{1}{\nu}\, \delta[x(t) - a] \mid x'(t) \mid S_b[x'(t)] \\
&= \frac{1}{\nu} \int_{-\infty}^{\infty} dx \int_{-\infty}^{b} d\xi \delta(x - a) \mid \xi \mid p(\xi, x) \\
&= \frac{1}{\nu} \int_{-\infty}^{b} d\xi \mid \xi \mid p(\xi, a).
\end{aligned}$$

Here we have appealed to the ergodic theorem. The empirical density for $\xi$ then follows by differentiating with respect to $b$,

$$p(\xi \mid x(0) = a)_{\text{emp}} = \frac{1}{\nu} \mid \xi \mid p(\xi, a). \tag{2.2}$$

Now the denominator of (2.1) is the probability that $x(t) - a$ have a zero in the small time interval $\delta$. Evaluating the integrals, one finds this probability to be $\sqrt{2\alpha/\pi}p(a)\delta$ to first order in $\delta$. It follows then that $\nu = \sqrt{2\alpha/\pi}p(a)$. Inserting this value in (2.2) yields

$$p(\xi \mid x(0) = a)_{\text{emp}} = \frac{\mid \xi \mid}{2\alpha}\, e^{-\frac{\xi^2}{2\alpha}} = p(\xi \mid x(0) = a)_{\text{h.w.}}.$$

It might be mentioned that the interpretation of conditional probabilities in the h.w. sense is intimately connected with the definition of the mean recurrence time as introduced into statistical physics by Smoluchowski.

Let $x(t)$ be an ergodic process and consider the discrete observations $x(0)$, $x(\tau)$, $x(2\tau)$, $\cdots$. For these observations the mean recurrence time of the state defined by $|x(t)| < \delta$ is given by Smoluchowski's formula

$$\theta_{\delta,\tau} = \tau \frac{1 - \{\Pr\{|x(0)| < \delta\}}{\Pr\{|x(0)| < \delta, |x(\tau)| > \delta\}}$$

which he derived by direct time average considerations. (For a derivation of this formula as well as a discussion of its connection with the ergodic theorem see [3].) Denoting by $W(x)$ the probability density of $x(t)$ and by $W_\tau(x, y)$ the joint probability density of $x(t)$ and $x(t + \tau)$, we get

$$\theta_{\delta,\tau} = \tau \frac{1 - \int_{-\delta}^{\delta} W(x)\, dx}{\displaystyle\iint\limits_{\substack{|x|<\delta \\ |y|>\delta}} W_\tau(x, y)\, dx\, dy}.$$

For a Gaussian ergodic process for which $x'(t)$ is defined, we can go further. Since

$$x(\tau) \sim x(0) + \tau x'(0)$$

and $x(0)$ and $x'(0)$ are independent, we have

$$\iint\limits_{\substack{|x|<\delta \\ |y|>\delta}} W_\tau(x, y)\, dx\, dy \sim \frac{1}{2\pi\sqrt{\alpha}} \iint\limits_{\substack{|x|<\delta \\ |x+\tau\xi|>\delta}} e^{-\left(\frac{x^2}{2}+\frac{\xi^2}{2\alpha}\right)}\, dx\, d\xi$$

where $\alpha = E[x'(0)]^2$. Now

$$\iint\limits_{\substack{|x|<\delta \\ |x+\tau\xi|>\delta}} e^{-\left(\frac{x^2}{2}+\frac{\xi^2}{2\alpha}\right)}\, dx\, d\xi = \int_{-\delta}^{\delta} dx e^{-\frac{x^2}{2}} \int_{\frac{\delta-x}{\tau}}^{\infty} d\xi e^{-\frac{\xi^2}{2\alpha}} + \int_{-\delta}^{\delta} dx e^{-\frac{x^2}{2}} \int_{-\infty}^{\frac{-\delta-x}{\tau}} d\xi e^{-\frac{\xi^2}{2\alpha}}.$$

In the first of these integrals set $x = \delta - y\tau$. In the second, put $x = -\delta - y\tau$. There results

$$\tau \int_0^{\frac{2\delta}{\tau}} dy e^{-\frac{(\delta-y\tau)^2}{2}} \int_y^{\infty} d\xi e^{-\frac{\xi^2}{2\alpha}} + \tau \int_{-\frac{2\delta}{\tau}}^{0} dy e^{-\frac{(\delta+y\tau)^2}{2}} \int_{-\infty}^{y} d\xi e^{-\frac{\xi^2}{2\alpha}}$$

and hence

$$\lim_{\delta\to 0} \lim_{\tau\to 0} \theta_{\delta,\tau} = \frac{1}{\frac{1}{\pi\sqrt{\alpha}} \int_0^{\infty} dy \int_y^{\infty} d\xi e^{-\frac{\xi^2}{2\alpha}}} = \frac{\pi\sqrt{\alpha}}{\int_0^{\infty} y e^{-\frac{y^2}{2\alpha}}\, dy}$$

which agrees with the known result of Rice for the mean distance between zeros.

**3. Joint distribution for large positive excursions.** Let $x(t)$ be a continuous parameter ergodic differentiable Gaussian process with mean zero and covari-

ance function $\rho(\tau)$. For convenience we choose $\rho(0) = 1$ and assume that in some interval about $\tau = 0$,

$$\rho(\tau) = 1 - \frac{\alpha}{2}\tau^2 + o(\tau^2), \qquad \alpha > 0. \tag{3.1}$$

Let $\theta = \theta(a)$ denote the expected length of the intervals during which $x(t) \geqq a$. Then [2]

$$\theta = \frac{\pi e^{\frac{a^2}{2}}}{\sqrt{\alpha}}\left[1 - \sqrt{\frac{2}{\pi}}\int_0^a e^{-\frac{t^2}{2}}\,dt\right] \tag{3.2}$$

and

$$\theta \sim \sqrt{\frac{2\pi}{\alpha}}\frac{1}{a} \tag{3.3}$$

for large positive $a$.

In this and the next several sections, we study some limiting properties of the related process

$$\Delta(t, \theta) = \frac{x(\theta t) - a}{\theta} \tag{3.4}$$

as $a \to +\infty$, (or, equivalently, as $\theta \to 0$ through positive values). We shall generally be concerned only with properties of $\Delta(t, \theta)$ conditioned by

$$\Delta'(0, \theta) = \left.\frac{\partial \Delta}{\partial t}\right|_{t=0} \geqq 0 \quad \text{and} \quad \Delta(0, \theta) = 0$$

in either the h.w. or v.w. sense of Section 2.

The main result of this section is that, as $a \to \infty$, the $n$-dimensional joint distribution function of $\Delta(t_1, \theta), \Delta(t_2, \theta), \cdots, \Delta(t_n, \theta)$ conditioned in the v.w. sense by $\Delta'(0, \theta) \geqq 0$, $\Delta(0, \theta) = 0$ approaches the singular $n$-dimensional half-normal distribution function of the random variables

$$\Delta_i = -\sqrt{\frac{\alpha\pi}{2}}\,t_i^2 + \sqrt{\alpha}\,t_i\,\xi, \qquad i = 1, \cdots, n, \tag{3.5}$$

where $\xi$ is a random variable with probability density

$$p(\xi) = \begin{cases} 0, & \xi < 0 \\ \sqrt{\frac{2}{\pi}}\,e^{-\frac{\xi^2}{2}}, & \xi \geqq 0. \end{cases} \tag{3.6}$$

If the conditioning is done in the h.w. sense, the result remains the same except that $\xi$ now has the Rayleigh density

$$p(\xi) = \begin{cases} 0, & \xi < 0 \\ \xi e^{-\frac{\xi^2}{2}}, & \xi \geqq 0. \end{cases} \tag{3.7}$$

In one sense, then, as $a \to +\infty$, the sample functions of the conditioned $\Delta(t, \theta)$ process become a family of random parabolas,

$$\Delta = -\sqrt{\frac{\alpha\pi}{2}}\, t^2 + \sqrt{\alpha}\, t\xi, \tag{3.8}$$

where $\xi$ has either a half-normal or Rayleigh distribution according as $\Delta(0, \theta)$ is conditioned to be zero in the v.w. or h.w. sense. In terms of the original $x(t)$ process, one can say that, when properly scaled and normalized, the excursions of $x(t)$ above the level $a$ approach parabolas as $a \to +\infty$.

It is worth noting that these results require only the existence of the first derivative of $x(t)$. Processes with pathologies only in higher order derivatives, such as the harmonic oscillator of physics, are sufficiently "tamed" by the normalization and scaling indicated in (3.4) to give the limits mentioned.

We obtain the limiting conditional distribution function for $\Delta(t_1, \theta), \cdots, \Delta(t_n, \theta)$ by computing the characteristic function, $\varphi_a(\eta_1, \eta_2, \cdots, \eta_n)$, for these quantities and determining the limiting function

$$\varphi(\eta_1, \cdots, \eta_n) = \lim_{a\to+\infty} \varphi_a(\eta_1, \cdots, \eta_n).$$

By a well-known theorem ([4], p. 102), $\varphi(\eta_1, \cdots, \eta_n)$, if continuous at $\eta_1 = \eta_2 = \cdots = \eta_n = 0$, is the characteristic function of the limiting distribution function for $\Delta(t_1, \theta), \cdots, \Delta(t_n, \theta)$.

Let $\xi = x'(0)$, $x_i = x(\theta t_i)$, $i = 0, 1, 2, \cdots, n$, and $t_0 = 0$. Then

$$\begin{aligned} p(x_1, \cdots, x_n \mid \xi \geqq 0, x_0 = a)_{\text{v.w.}} &\\ &= \frac{\int_0^\infty d\xi p(\xi, x_0, x_1, \cdots, x_n)\big|_{x_0=a}}{\frac{1}{2}p(x_0)\big|_{x_0=a}} \\ &= \frac{\int_0^\infty d\xi p(x_1, \cdots, x_n \mid \xi, x_0 = a)p(\xi, x_0)\big|_{x_0=a}}{\frac{1}{2}p(x_0)\big|_{x_0=a}} \\ &= 2\int_0^\infty d\xi p(x_1, \cdots, x_n \mid \xi, x_0 = a)p(\xi). \end{aligned} \tag{3.9}$$

Now $p(x_1, \cdots, x_n|\xi, x_0 = a)$ is an $n$-variate Gaussian density (see Appendix). The conditional means and covariances are readily computed:

$$m_i = E(x_i \mid \xi, x_0 = a) = \rho(\theta t_i)a - \frac{1}{\alpha}\rho'(\theta t_i)\xi \tag{3.10}$$

and

$$\begin{aligned} \lambda_{ij} &= E[(x_i - m_i)(x_j - m_j) \mid \xi, x_0 = a] \\ &= \rho[\theta(t_i - t_j)] - \rho(\theta t_i)\rho(\theta t_j) - \frac{1}{\alpha}\rho'(\theta t_i)\rho'(\theta t_j), \\ &\qquad\qquad i, j, = 1, 2, \cdots, n. \end{aligned} \tag{3.11}$$

One can write, therefore,

$$p(x_1, \cdots, x_n \mid \xi, x_0 = a) = (2\pi)^{-n} \int_{\infty}^{\infty} d\eta_1 \cdots \int_{\infty}^{\infty} d\eta_n \, e^{-i \sum \eta_j (x_j - m_j) - \frac{1}{2} \sum \lambda_{jk} \eta_j \eta_k}.$$

Introduce this expression into (3.9), substitute $x_i = a + \theta\Delta_i$ and multiply the entire expression by $\theta^n$ to obtain the conditional density function for $\Delta(t_1, \theta), \cdots, \Delta(t_n, \theta)$ in the form

$$p[\Delta_1, \cdots, \Delta_n \mid \Delta'(0, \theta) \geqq 0, \Delta(0, \theta) = 0]_{\text{v.w.}}$$

$$= \frac{2\theta^n}{(2\pi)^n} \int_0^{\infty} d\xi \int_{\infty}^{\infty} d\eta_1 \cdots \int_{\infty}^{\infty} d\eta_n \, e^{-i \sum \eta_j (a + \theta\Delta_j - m_j) - \frac{1}{2} \sum \lambda_{jk} \eta_j \eta_k} \frac{e^{-\frac{\xi^2}{2\alpha}}}{\sqrt{2\pi\alpha}}.$$

Let $\xi = \sqrt{\alpha}\xi'$, $\theta\eta_i = \eta_i'$, $i = 1, \cdots, n$. Introduce the value of $m_i$ given in (3.10), interchange the order of integration which is a step easily justified, and omit the primes. There results

$$p[\Delta_1, \cdots, \Delta_n \mid \Delta'(0, \theta) \geqq 0, \Delta(0, \theta) = 0]_{\text{v.w.}}$$

$$= \frac{1}{(2\pi)^n} \int_{\infty}^{\infty} d\eta_1 \cdots \int_{\infty}^{\infty} d\eta_n \, e^{-i \sum \eta_j \Delta_j} \varphi_a(\eta_1, \cdots, \eta_n)$$

where

$$\varphi_a(\eta_1, \cdots, \eta_n) = \exp\left[-i \sum \eta_j \frac{a}{\theta} [1 - \rho(\theta t_j)] - \tfrac{1}{2} \sum \frac{\lambda_{jk}}{\theta^2} \eta_j \eta_k\right]$$

$$\cdot \int_0^{\infty} d\xi \sqrt{\frac{2}{\pi}} \, e^{-i \frac{\xi}{\alpha} \sum \eta_j \frac{\rho'(\theta t_j)}{\theta} - \frac{1}{2}\xi^2}.$$

On using (3.1), (3.3) and (3.11), one finds,

$$\varphi(\eta_1, \cdots, \eta_n) = \lim_{a \to \infty} \varphi_a(\eta_1, \cdots, \eta_n) = e^{-i\sqrt{\frac{\pi\alpha}{2}} \sum \eta_j t_j^2} \int_0^{\infty} d\xi \sqrt{\frac{2}{\pi}} \, e^{i\sqrt{\alpha}\xi \sum \eta_j t_j - \frac{1}{2}\xi^2}.$$

But this expression, which is continuous at $\eta_1 = \cdots = \eta_n = 0$, is just the characteristic function associated with the random variables (3.5), (3.6), as a trivial computation shows.

The determination of the limiting form of the joint distribution function for $\Delta(t_1, \theta), \ldots, \Delta(t_n, \theta)$ conditioned in the h.w. sense by $\Delta'(0, \theta) \geqq 0$, $\Delta(0, \theta) = 0$ proceeds in a similar manner. Here (3.9) is replaced by

$$p(x_1, \cdots, x_n \mid \xi \geqq 0, x_0 = a)_{\text{h.w.}}$$

$$= \frac{\int_0^{\infty} d\xi \xi p(\xi, x_0, x_1, \cdots, x_n) \mid_{x_0 = a}}{\int_0^{\infty} d\xi \xi p(\xi, x_0) \mid_{x_0 = a}}$$

$$= \sqrt{\frac{2\pi}{\alpha}} \int_0^{\infty} d\xi \xi p(x_1, \cdots, x_n \mid \xi, x_0 = a) p(\xi).$$

The remaining steps are as in the previous demonstration.

**4. Asymptotic distribution of first return time to positive level.** We now assume that

$$\rho(\tau) = 1 - \frac{\alpha}{2}\tau^2 + \frac{c_3}{3!}|\tau|^3 + \frac{c_4}{4!}\tau^4 + o(\tau^4). \tag{4.1}$$

Let $P_a(T)$ be the probability that $\Delta(t, \theta)$ be non-negative for $0 \leqq t \leqq T$ given that $\Delta'(0, \theta) \geqq 0$ and $\Delta(0, \theta) = 0$ in the h.w. sense. Then $Q_a(T) = -(dP_a/dT)$ is the probability density for the duration of the positive excursions of the $\Delta(t, \theta)$ process conditioned in the h.w. sense. In this section we show that

$$Q(T) \equiv \lim_{a\to\infty} Q_a(T) = \begin{cases} \dfrac{\pi T}{2} e^{-\frac{\pi}{4}T^2}, & T \geqq 0 \\ 0, & T < 0 \end{cases} \tag{4.2}$$

and

$$P(T) \equiv \lim_{a\to\infty} P_a(T) = \begin{cases} e^{-\frac{\pi}{4}T^2}, & T \geqq 0 \\ 0 \quad , & T < 0. \end{cases} \tag{4.3}$$

If the conditioning is done in the v.w. sense, the corresponding results are

$$Q(T) = \begin{cases} e^{-\frac{\pi}{4}T^2}, & T \geqq 0 \\ 0 \quad , & T < 0 \end{cases} \tag{4.4}$$

$$P(T) = \begin{cases} 1 - \displaystyle\int_0^T e^{-\frac{\pi}{4}x^2}\, dx, & T \geqq 0 \\ 0 \quad , & T < 0. \end{cases} \tag{4.5}$$

These results are consistent with the interpretation of the sample functions of the limiting $\Delta$ process as the family of random parabolas (3.8) with $\xi$ distributed according to (3.6) or (3.7). Note that the results (4.2)–(4.5) are independent of the parameters defining $\rho(\tau)$. All differentiable ergodic Gaussian processes, when scaled as here, have the same asymptotic distribution for the duration of excursions above a level.

To compute $P_a(T)$, we make use of the method of "inclusion and exclusion" [5], p. 89, in a manner analogous to that of Rice [6], p. 70. Let $A_i$ be the event "$x(t)$ assumes the value $a$ for some value of $t$ such that

$$i(T/n) \leqq t < (i + 1)(T/n)$$

given that $x'(0) \geqq 0$ and $x(0) = a$ in the h.w. sense." Then the probability, $W_a(T)$, that $x(t)$ be not less than $a$ for $0 \leqq t \leqq T$ given that $x'(0) \geqq 0$ and $x(0) = a$ in the h.w. sense is

$$W_a(T) = 1 - \sum_i \Pr[A_i] + \sum_{i<j} \Pr[A_i \cap A_j] \sum_{i<j<k} \Pr[A_i \cap A_j \cap A_k] + \cdots$$

In the limit as $n \to \infty$, this becomes

$$W_a(T) = 1 - \frac{1}{1!}\int_0^T dt_1\, p_1(t_1) + \frac{1}{2!}\int_0^T dt_1 \int_0^T dt_2\, p_2(t_1, t_2) - \frac{1}{3!}\int_0^T dt_1 \int_0^T dt_2 \int_0^T dt_3\, p_3(t_1, t_2, t_3) + \cdots$$

where

$$p_j(t_1, \cdots, t_j)\, dt_1 \cdots dt_j$$

is the probability that $x(t)$ assumes the value $a$ in each of the intervals $(t_1, t_1 + dt_1), \cdots, (t_j, t_j + dt_j)$ given that $x'(0) \geqq 0$ and $x(0) = a$ in the h.w. sense.

One has then, since $P_a(T) = W_a(\theta T)$,

$$P_a(T) = 1 - \frac{\theta}{1!}\int_0^T dt_1\, p_1(\theta t_1) + \frac{\theta^2}{2!}\int_0^T dt_1 \int_0^T dt_2\, p_2(\theta t_1, \theta t_2) - \cdots \tag{4.6}$$

and by differentiation

$$Q_a(T) = \theta p_1(\theta T) - \frac{\theta^2}{1!}\int_0^T dt_1\, p_2(\theta t_1, \theta T) + \cdots. \tag{4.7}$$

Here

$$p_n(\theta t_1, \cdots, \theta t_n) = \frac{\int_0^\infty d\xi_0 \int_{-\infty}^\infty d\xi_1 \cdots \int_{-\infty}^\infty d\xi_n\, \xi_0 \,|\xi_1| \cdots |\xi_n|\, p(\xi_0, \cdots, \xi_n, x_0, \cdots, x_n)\,|_{x'_i=a}}{\sqrt{\dfrac{\alpha}{2\pi}}\, p(x_0)\,|_{x_0=a}} \tag{4.8}$$

where

$$x_i = x(\theta t_i), \qquad \xi_i = x'(\theta t_i), \qquad t_0 = 0, \qquad i = 0,\ 1, \cdots, n$$

and $p$ denotes the joint density of the random variables indicated.

From the derivation of the method of inclusion and exclusion, the successive partial sums of (4.6) and (4.7) alternately over-estimate and under-estimate the limit sum. Therefore

$$0 \leqq P_a(T) - 1 + \theta \int_0^T dt_1\, p_1(\theta t_1) \leqq \frac{\theta^2}{2}\int_0^T dt_1 \int_0^T dt_2\, p_2(\theta t_1, \theta t_2), \tag{4.9}$$

$$0 \leqq Q_a(T) - \theta p_1(\theta T) \leqq \frac{\theta^2}{2}\int_0^T dt_1\, p_2(\theta t_1, \theta T). \tag{4.10}$$

We establish (4.2) and (4.3) by evaluating $\lim_{a\to\infty} \theta p_1(\theta t_1)$ and by showing the right members of (4.9) and (4.10) approach zero as $a \to \infty$. A completely analogous procedure gives (4.4) and (4.5).

To investigate the behavior of $\theta p_1(\theta t_1)$ for large $a$, we write (4.8) for $n = 1$ as

$$\theta\, p_1(\theta t_1) = \sqrt{\frac{2\pi}{\alpha}}\, \theta\, p(x_1 \mid x_0 = a)\, |_{x_1 = a}\, I_a\,, \tag{4.11}$$

$$I_a = \int_0^\infty d\xi_0 \int_{-\infty}^\infty d\xi_1\, \xi_0 \mid \xi_1 \mid p(\xi_0\,, \xi_1 \mid x_0 = a, x_1 = a). \tag{4.12}$$

An elementary calculation shows that

$$\theta \sqrt{\frac{2\pi}{\alpha}}\, p(x_1 \mid x_0 = a)_{x_1 = a} = \left[\frac{\theta^2}{\alpha(1 - \rho^2)}\right]^{\frac{1}{2}} e^{-\frac{a^2}{2}\frac{1-\rho}{1+\rho}}$$

where $\rho = \rho(\theta t_1)$. Using (4.1) and (3.3), there results

$$\lim_{a \to \infty} \theta \sqrt{\frac{2\pi}{\alpha}}\, p(x_1 \mid x_0 = a)\, |_{x_1 = a} = \frac{e^{-\frac{\pi}{4} t_1^2}}{\alpha t_1}\,. \tag{4.13}$$

The factor $p(\xi_0\,, \xi_1 | x_0 = a, x_1 = a)$ in (4.12) is a bivariate Gaussian density. The conditional means are found to be (see Appendix)

$$m_0 = E(\xi_0 \mid x_0 = x_1 = a) = -\frac{a\, \rho'(\theta t_1)}{1 + \rho(\theta t_1)} = -E(\xi_1 \mid x_0 = x_1 = a) = -m_1\,.$$

The conditional covariances are

$$\lambda_{00} = E[(\xi_0 - m_0)^2 \mid x_0 = x_1 = a] = \frac{\alpha - [\rho'(\theta t_1)]^2 - \alpha\rho^2(\theta t_1)}{1 - \rho^2(\theta t_1)}$$

$$= \lambda_{11} = E[(\xi_1 - m_1)^2 \mid x_0 = x_1 = a],$$

$$\lambda_{01} = \frac{\rho^2(\theta t_1)\rho''(\theta t_1) - \rho''(\theta t_1) - \rho(\theta t_1)[\rho'(\theta t_1)]^2}{1 - \rho^2(\theta t_1)}\,.$$

As $a \to \infty$, $m_0 \to \sqrt{\alpha\pi/2}\, t_1$, $m_1 \to -\sqrt{\alpha\pi/2}\, t_1$; the covariances are $O(\theta)$ if $c_3 \neq 0$ and $o(\theta)$ if $c_3 = 0$. By standard arguments, then, as $a \to \infty$ the contribution to $I_a$ comes entirely from the neighborhood of the point $(m_0\,, m_1)$ and

$$\lim_{a \to \infty} I_a = \begin{cases} \dfrac{\alpha\pi}{2}\, t_1^2, & t_1 \geqq 0 \\ 0\quad, & t_1 < 0. \end{cases}$$

Combining this result with (4.13), we find

$$\lim_{a \to \infty} \theta p_1(\theta t_1) = \begin{cases} \dfrac{\pi}{2}\, t_1\, e^{-\frac{\pi}{4} t_1^2}, & t_1 \geqq 0 \\ 0\quad, & t_1 < 0. \end{cases} \tag{4.14}$$

To show that the right member of the (4.10) approaches zero, we write (4.8) for $n = 2$ in the form

$$\theta^2 p_2(\theta t_1\,, \theta t_2) = p(x_1\,, x_2 \mid x_0 = a)\, |_{\substack{x_1 = a \\ x_2 = a}} \cdot J_a \tag{4.15}$$

with

$$J_a = \theta^2 \sqrt{\frac{2\pi}{\alpha}}$$

(4.16)

$$\cdot \int_0^\infty d\xi_0 \int_{-\infty}^\infty d\xi_1 \int_{-\infty}^\infty d\xi_2\, \xi_0 \mid \xi_1 \mid \mid \xi_2 \mid p(\xi_0, \xi_1, \xi_2 \mid x_0 = x_1 = x_2 = a).$$

The first factor of (4.15) is of the form

$$p(x_1, x_2 \mid x_0 = a) \mid_{\substack{x_1=a\\x_2=a}} = \frac{e^{-\frac{a^2}{2}h(\theta)}}{2\pi\sqrt{d}}$$

where

$$d = 1 + 2\rho(\theta t_1)\rho(\theta t_2)\rho[\theta(t_2 - t_1)] - \rho^2(\theta t_1) - \rho^2(\theta t_2) - \rho^2[\theta(t_2 - t_1)]$$

and

$$h(\theta) = 2\,\frac{[1 - \rho(\theta t_1)][1 - \rho(\theta t_2)]\,\{1 - \rho[\theta(t_2 - t_1)]\}}{d}.$$

A lengthy calculation shows that as $a \to \infty$,

(4.17)
$$d = \begin{cases} \frac{2}{3}\alpha c_3\, t_1^2 t_2 (t_2 - t_1)^2 \theta^5 + o(\theta^5), & c_3 \neq 0 \\ \frac{1}{4}\alpha(c_4 - \alpha^2) t_1^2 t_2^2 (t_2 - t_1)^2 \theta^6 + o(\theta^6), & c_3 = 0, \end{cases}$$

so that

(4.18)
$$\frac{a^2}{2} h(\theta) = \begin{cases} \frac{3\pi\alpha t_2}{8c_3}\frac{1}{\theta} + o\left(\frac{1}{\theta}\right), & c_3 \neq 0 \\ \frac{\pi\alpha}{c_4 - \alpha^2}\frac{1}{\theta^2} + o\left(\frac{1}{\theta^2}\right), & c_3 = 0. \end{cases}$$

The first factor of (4.15) therefore approaches zero at least as fast as $A\theta^{3/2}e^{-B/\theta}$.

The proof is completed by showing that $J_a$ is $O(1/\theta^r)$ for some finite $r$ so that from (4.15) $\theta^2 p_2(\theta t_1, \theta t_2) \to 0$ as $a \to \infty$. Now

(4.19)
$$\begin{aligned} J_a/\theta^2 \sqrt{\frac{2\pi}{\alpha}} &\leqq \int_{-\infty}^\infty d\xi_0 \int_{-\infty}^\infty d\xi_1 \int_{-\infty}^\infty d\xi_2 \mid \xi_0 \mid \mid \xi_1 \mid \mid \xi_2 \mid p(\xi_0, \xi_1, \xi_2 \mid x_0 = x_1 = x_2 = a) \\ &\leqq 8 + E(\xi_0^2 \xi_1^2 \xi_2^2 \mid x_0 = x_1 = x_2 = a). \end{aligned}$$

This conditional expectation, however, is a multi-nomial in the conditional means and variances of the $\xi$'s. These latter quantities in turn are rational functions of $a$, $\rho(\theta t_1)$, $\rho(\theta t_2)$, $p[\theta(t_2 - t_1)]$, $\rho'(\theta t_1)$, $\cdots$, $\rho''[\theta(t_2 - t_1)]$. It follows then that the right side of (4.19), and hence $J_a$ also, is $O(1/\theta^r)$.

We note in passing that in the case $c_3 = 0$ the factor $c_4 - \alpha^2$ in (4.17) and (4.18) is non-negative and vanishes only if $\rho(\tau) = \cos\beta\tau$. This can be established as follows. From (4.1) it is easily seen that when $c_3 = 0$, $\rho''(0)$ and $\rho^{(4)}(0)$ exist and are given by

$$-\alpha = \rho''(0) \qquad c_4 = \rho^{(4)}(0). \tag{4.20}$$

Now since $\rho(\tau)$ is a covariance function, we can write

$$\rho(\tau) = \int_{-\infty}^{\infty} e^{2\pi i\lambda\tau}\,dF(\lambda)$$

where $F(\lambda)$ is non-decreasing. It follows then (see [4] p. 90 for a similar argument involving a distribution function and its characteristic function) that the second and fourth moments of $F$ exist and that

$$\rho''(0) = -\alpha = -4\pi^2 \int_{-\infty}^{\infty} \lambda^2\,dF(\lambda),$$

$$\rho^{(4)}(0) = c_4 = 16\pi^4 \int_{-\infty}^{\infty} \lambda^4\,dF(\lambda).$$

The Schwartz inequality then gives

$$c_4 \geqq \alpha^2$$

with equality only if $\rho(\tau)$ is of the form $\cos\beta\tau$. Our derivation of (4.2)–(4.5) fails in this case. Indeed, we have already excluded this process with covariance $\cos\beta\tau$ from consideration since it is not ergodic. The results (4.2)–(4.5) are still valid for this process, however, as a separate calculation, omitted here, shows.

**5. Asymptotic distribution of first return time to negative levels.** As in the preceding section, let $Q_a(T)$ be the probability density for the duration of the excursions above the value $a$ of the $\Delta(t, \theta)$ process conditioned in the h.w. sense. If in addition to (4.1), $\rho(\tau)$ and its first two derivatives approach zero as $\tau \to \infty$, then

$$\lim_{a\to-\infty} Q_a(T) = 2e^{-2T}.$$

This result follows readily from (4.7) and (4.8) and the asymptotic formula for $\theta$ for large negative values of $a$,

$$\theta \sim \frac{2\pi}{\sqrt{\alpha}}\, e^{\frac{a^2}{2}},$$

obtained from (3.2). For large negative $a$, the random variables $x_0 = x(\theta t_1), \cdots, x_n = x(\theta t_n)$, $\xi_0 = x'(\theta t_0), \cdots, \xi_n = n'(\theta t_n)$ tend toward independence and the density in the numerator of (4.8) approaches

$$\frac{e^{-\frac{1}{2\alpha}\Sigma\xi_i^2-\frac{(n+1)}{2}a^2}}{(2\pi)^{n+1}\dfrac{n+1}{\alpha^2}}$$

in a uniformly continuous way. One finds, then, $\theta^n p_n(\theta t_1, \cdots, \theta t_n) \to 2^n$ and the series (4.7) sums to $2e^{-2r}$.

If the conditioning is done in the v.w. sense, the same result is found. It is to be noted that this limiting distribution, like those obtained for large positive $a$, (4.2)–(4.5), is independent of the covariance $\rho(\tau)$.

## APPENDIX

The detailed calculations of this paper make frequent use of the multivariate conditional densities for Gaussian variables. Since these densities do not appear to be readily available in the literature, we present them here for the reader's convenience. They can be derived with a little effort from material given in many texts, e.g. [4] or [7], pp. 27–30.

Let $\xi_1, \cdots, \xi_n$ be jointly Gaussian with $E\xi_i = 0, E\xi_i\xi_j = \lambda_{ij}, i, j = 1, 2, \cdots, n$. Then

$$p(\xi_{p+1}, \cdots, \xi_n \mid \xi_1, \cdots, \xi_p) = \frac{e^{-\frac{1}{2}\sum_{p+1}^{n} \mu_{ij}^{-1}(\xi_i - m_i)(\xi_j - m_j)}}{(2\pi)^{\frac{n-p}{2}} \mid \mu \mid^{\frac{1}{2}}}$$

where

$$m_i = E(\xi_i|\xi_1, \cdots, \xi_p) = \sum_{j=1}^{p} \beta_{ij}\xi_j, \qquad i = p + 1, \cdots, n$$

$$\mu_{ij} = E[(\xi_j - m_i)(\xi_j - m_j)|\xi_1, \cdots, \xi_p], \quad i, j = p + 1, \cdots, n$$

and

$$\beta_{ij} = \frac{1}{d}\begin{vmatrix} \lambda_{11} & \cdots & \lambda_{1(j-1)} & \lambda_{1i} & \lambda_{1(j+1)} & \cdots & \lambda_{1p} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \lambda_{p1} & \cdots & \lambda_{p(j-1)} & \lambda_{pi} & \lambda_{p(j+1)} & \cdots & \lambda_{pp} \end{vmatrix}$$

$$\mu_{ij} = \frac{1}{d}\begin{vmatrix} \lambda_{ij} & \lambda_{i1} & \cdots & \lambda_{ip} \\ \lambda_{1j} & \lambda_{11} & \cdots & \lambda_{1p} \\ \vdots & \vdots & & \vdots \\ \lambda_{pj} & \lambda_{p1} & \cdots & \lambda_{pp} \end{vmatrix}$$

$$d = \begin{vmatrix} \lambda_{11} & \cdots & \lambda_{1p} \\ \vdots & & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pp} \end{vmatrix}.$$

## REFERENCES


[1] D. S. Palmer, "Properties of random functions," *Proc. Camb. Phil. Soc.*, Vol. 52 (1956), pp. 672–686.

[2] S. O. Rice, "Duration of fades in radio transmission," *Bell System Tech. J.*, Vol. 37 (1958), pp. 581–635.

[3] M. Kac, "On the notion of recurrence in discrete stochastic processes," *Bull. Am. Math. Soc.*, Vol. 53 (1947), pp. 1002–1010.

[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.

[5] W. Feller, *Probability Theory*, John Wiley & Sons, New York, 1957.

[6] S. O. Rice, "Mathematical analysis of random noise," Reprinted in *Noise and Stochastic Processes*, Dover Publications, New York, (1954). Original *Bell System Tech. J.*, Vol. 24 (1945), pp. 46–156.

[7] T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 1958.

# THE CAPACITY OF A CLASS OF CHANNELS[1]

By David Blackwell, Leo Breiman, and A. J. Thomasian

*University of California, Berkeley*

**1. Summary.** Shannon's basic theorem on the capacity of a channel is generalized to the case of a class of memoryless channels. A generalized capacity is defined and is shown to be the supremum of attainable transmission rates when the coding and decoding procedure must be satisfactory for every channel in the class.

**2. Definitions and Introduction.** For any positive integer $n$ and any set $\mathfrak{A}$ we denote by $\mathfrak{A}^{(n)}$ the set of all $n$-tuples $(x_1, \cdots, x_n)$ with each $x_i \varepsilon \mathfrak{A}$.

A channel, denoted by $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ or by $P(y \mid x)$, consists of two finite sets $\mathfrak{A}$, $\mathfrak{B}$ having $a \geqq 2$, $b \geqq 2$ elements, respectively, and a set of probability distributions $P(\cdot \mid x)$ on $\mathfrak{B}$, one for each $x \varepsilon \mathfrak{A}$. $P(y \mid x)$ is interpreted as the probability of receiving $y \varepsilon \mathfrak{B}$ given that $x \varepsilon \mathfrak{A}$ was transmitted.

The $n$-extension of a channel $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ is the channel $(\mathfrak{A}^{(n)}, \mathfrak{B}^{(n)}, P(v \mid u))$ where $v = (y_1, \cdots, y_n) \varepsilon \mathfrak{B}^{(n)}$, $u = (x_1, \cdots, x_n) \varepsilon \mathfrak{A}^{(n)}$ and $P(v \mid u) = \prod_{i=1}^{n} P(y_i \mid x_i)$.

When considering a class of channels, $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y \mid x))$ for $\gamma \varepsilon \mathfrak{C}$, where $\mathfrak{C}$ is an index set, we shall always assume that the $\mathfrak{A}$, $\mathfrak{B}$ sets are the same for each channel in the class. We shall sometimes denote such a class of channels by $\mathfrak{C}$, the index set.

A $(G, \epsilon_n, n)$ code for a class $\mathfrak{C}$ of channels for $G \geqq 1$, $\epsilon_n \geqq 0$, $n$ a positive integer, is a sequence of $[G]$ distinct elements of $\mathfrak{A}^{(n)}$; $u_1, \cdots, u_{[G]}$; where $[G]$ is the largest integer $\leqq G$, and a sequence of $[G]$ disjoint subsets of $\mathfrak{B}^{(n)}$; $B_1, \cdots, B_{[G]}$; such that

$$P_\gamma(B_i^c \mid u_i) \leqq \epsilon_n \quad \text{for} \quad i = 1, \cdots, [G] \quad \text{and all} \quad \gamma \varepsilon \mathfrak{C}.$$

The set $\{u_1, \cdots, u_{[G]}\}$ is called the set of input messages of the code and $B_i$ is called the decoding set for $u_i$. We think of an input letter $u_i$ of the code as being selected arbitrarily and transmitted over an unknown one of the channels $P_\gamma$, $\gamma \varepsilon \mathfrak{C}$. The letter $v$ is received with probability $P_\gamma(v \mid u)$ and if $v \varepsilon B_j$ it is decoded as $u_j$. Thus, the probability is $\leqq \epsilon_n$ that any input message $u_i$ will be transmitted so as to be not decoded as $u_i$; regardless of which channel in the class $\mathfrak{C}$ is used.

An $R \geqq 0$ is an attainable transmission rate for a class $\mathfrak{C}$ of channels if there exists a sequence of $(e^{Rn}, \epsilon_n, n)$ codes for $\mathfrak{C}$ with $\epsilon_n \to 0$. Since $\mathfrak{A}^{(n)}$ has only $a^n$ points we know that any attainable rate $R \leqq \log a$. Clearly 0 is an attainable rate for any class of channels. For any class of channels $\mathfrak{C}$ we define $T = T(\mathfrak{C})$ to be the supremum of the set of attainable rates for $\mathfrak{C}$.

Received February 16, 1959.

[1] This research was supported by the Office of Naval Research under Contract Nonr-222(53).

If $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y \mid x))$ for $\gamma \varepsilon \mathfrak{C}$ is a class of channels and $Q(x)$ is a given probability distribution on $\mathfrak{A}$ then for each $\gamma \varepsilon \mathfrak{C}$ we let $P_\gamma(x, y) = P_\gamma(y \mid x)Q(x)$ and we define on $\mathfrak{A} \times \mathfrak{B}$ the random variable $J_\gamma$ by

$$J_\gamma(x, y) = \log \frac{P_\gamma(x, y)}{P_\gamma(x)P_\gamma(y)} \quad \text{if} \quad P_\gamma(x, y) > 0$$

$$= 0 \quad \text{if} \quad P_\gamma(x, y) = 0.$$

The dependence of $P_\gamma$ and $J_\gamma$ on $Q$ will usually not be exhibited. Since we will often be interested in expressions of the form $x \log x$ it is natural to define $\log 0 = 0$. We will denote the expectation of a random variable $X$ with respect to the $P_\gamma$ distribution by $E_\gamma X$. If $\mathfrak{C}$ has only one element we may drop the subscript $\gamma$. Finally for any class $\mathfrak{C}$ of channels we define the capacity of the class $\mathfrak{C}$ by

$$C(\mathfrak{C}) = C = \sup_{Q(x)} \inf_{\gamma \varepsilon \mathfrak{C}} E_\gamma J_\gamma$$

where the sup is over all distributions $Q$ on $\mathfrak{A}$.

In the case considered by Shannon, $\mathfrak{C}$ has only one element and our formula reduces to $C = \sup_Q EJ$, which is the usual formula for the capacity of a memoryless channel. Shannon's theorem then states that $T = C$. $T \geqq C$, $T \leqq C$ are called the direct and converse halves, respectively. This theorem for a single channel has been proved in various ways and under various conditions by Shannon [12], [13], McMillan [11], Feinstein [6], Khinchin [9], Wolfowitz [14], Blackwell, Breiman, and Thomasian [1]. We will show that within the framework that has been set up

$$T(\mathfrak{C}) = C(\mathfrak{C})$$

always holds true. This result follows immediately from Theorem 1 which also gives an exponential error bound for any rate $R < C$.

THEOREM 1: *Let $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y \mid x))$ for $\gamma \varepsilon \mathfrak{C}$ be any class of channels.*

(a) *For any integer $n$ and any $R > 0$ such that $0 \leqq C - R \leqq 1/2$ there is an $(e^{Rn}, \epsilon_n, n)$ code for $\mathfrak{C}$ with*

$$\epsilon_n = A e^{-\frac{(C-R)^2}{B} n}$$

*where*

$$A = \left[\frac{2^{10} ab^3}{(C - R)^2}\right]^{2ab} \quad \text{and} \quad B = 2^7 ab.$$

(b) *For any integer $n$ and $R > C$ if $e^{Rn} \geqq 2$ then any $(e^{Rn}, \epsilon_n, n)$ code for $\mathfrak{C}$ must satisfy*

$$\epsilon_n \geqq 1 - \frac{C + \frac{\log 2}{n}}{R - \frac{\log 2}{n}}.$$

The sequence of steps used in proving Theorem 1 will be outlined. Theorem 2 presents a basic inequality, for a single channel, which is contained implicitly

in Feinstein [8]. This inequality is of independent interest since it gives the same bound for the maximum probability of error that Shannon [13] gives for the average probability of error. Theorem 2 permits a simple proof of $T \geqq C$ for a single channel. Lemma 2 shows that $\sup_Q$ in the definition of $C(\mathcal{C})$ can be replaced by $\max_Q$. Theorem 3 gives an exponential bound on the error of a code for one channel, which depends only on $a$, $b$, $(C - R)^2$. This is convenient in that the particular probabilities $P(y \mid x)$ may not be known and, in any case, need not be computed with. Results related to Theorem 3 have been given by Elias [3] and [4], Feinstein [7], Shannon [13], and Wolfowitz [14].

Lemma 3 generalizes the inequality of Theorem 2 to the case when $\mathcal{C}$ has a finite number of elements, and Theorem 4 generalizes the exponential error bound of Theorem 3 to this case.

Lemma 4 shows that for a given $\mathfrak{A}$, $\mathfrak{B}$ there is a large finite number of channels on $\mathfrak{A}$, $\mathfrak{B}$ such that any channel on $\mathfrak{A}$, $\mathfrak{B}$ is close, in several senses, to one of them. Lemma 5 shows that if a channel has a sequence of codes $(e^{Rn}, \epsilon_n, n)$ with $\epsilon_n = e^{-Bn}$ for large $n$, with $B > 0$, then this same sequence of codes can be used for all channels in a certain neighborhood of the channel. This result justifies some of our attention to exponential error bounds. The technique of Lemma 5 can also be used to get some similar results when the channel probabilities vary from letter to letter.

At this point the direct half of Theorem 1 is demonstrated by approximating the class $\mathcal{C}$ of channels by a certain finite set of channels $\mathcal{C}'$ from Lemma 4; obtaining an exponential error bound code for $\mathcal{C}'$ from Theorem 4; and using Lemma 5 to show that such a code must be satisfactory for $\mathcal{C}$.

The converse half of Theorem 1 is then proved.

Before proceeding to the proofs we pause to clear up one point. It is obvious that

$$C(\mathcal{C}) \leqq \inf_{\gamma \in \mathcal{C}} \sup_{Q(x)} E_\gamma J_\gamma ,$$

i.e., $C(\mathcal{C}) \leqq$ the capacity of every channel in $\mathcal{C}$. We now exhibit an example where $C(\mathcal{C}) \neq$ inf of the capacities of channels in $\mathcal{C}$. Let $\mathfrak{A} = \mathfrak{B} = \{1, 2, 3, 4\}$, $\mathcal{C} = \{1, 2\}$, and let $P_1(y \mid x)$ and $P_2(y \mid x)$ be defined by the left and right following matrices, respectively.

$$\begin{pmatrix} \frac12 & \frac12 & 0 & 0 \\ 0 & 0 & \frac12 & \frac12 \\ \frac14 & \frac14 & \frac14 & \frac14 \\ \frac14 & \frac14 & \frac14 & \frac14 \end{pmatrix} \begin{pmatrix} \frac14 & \frac14 & \frac14 & \frac14 \\ \frac14 & \frac14 & \frac14 & \frac14 \\ \frac12 & \frac12 & 0 & 0 \\ 0 & 0 & \frac12 & \frac12 \end{pmatrix}$$

Let $Q(x)$ be any distribution on $\mathfrak{A}$ and let $H_i(Y) = -\sum_y P_i(y) \log P_i(y)$, $H_i(Y \mid X) = -\sum_x Q(x) \sum_y P_i(y \mid x) \log P_i(y \mid x)$. Using the fact that $\log x = (\log 2)\log_2 x$ we see that $(\log 2)^{-1} H_1(Y \mid X) = Q(1) + Q(2) + 2Q(3) + 2Q(4) = 1 + Q(3) + Q(4)$. Also from Feinstein [8], p. 15 we have $(\log 2)^{-1} H_1(Y) \leqq 2$ so that $E_1 J_1 = H_1(Y) - H_1(Y \mid X) \leqq (\log 2)(Q(1) + Q(2))$. Similarly $E_2 J_2 \leqq (\log 2)(Q(3) + Q(4))$ so that $C(\mathcal{C}) \leqq (1/2) \log 2$. The case $Q(i) = 1/4$ for $i = 1, \cdots, 4$ shows that $C(\mathcal{C}) = (1/2) \log 2$; the case

$Q(1) = Q(2) = 1/2$ shows the capacity of channel one to be log 2; the case $Q(3) = Q(4) = 1/2$ shows the capacity of channel two to be log 2. Thus for this example

$$\tfrac{1}{2}\log 2 = C(\mathfrak{C}) < \inf_{\gamma \epsilon \mathfrak{C}} \sup_{Q(x)} E_\gamma J_\gamma = \log 2.$$

**3. A basic inequality.**

THEOREM 2: *For any channel* $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$, *any distribution* $Q(x)$ *on* $\mathfrak{A}$, $\alpha > 0, G \geqq 1$ *there is a* $(G, \epsilon, 1)$ *code for the channel with* $\epsilon = Ge^{-\alpha} + P(J \leqq \alpha)$.

PROOF: It is clearly sufficient to construct an $(M, \epsilon, 1)$ code with the same $\epsilon$ as in the theorem and with $M \geqq G$. Let $A = [J > \alpha]$ and for any $x_0 \epsilon \mathfrak{A}$ let $A_{x_0} = \{(x, y) \mid (x_0, y) \epsilon A\}$. $P(J \leqq \alpha) \leqq \epsilon$ so that $P(A) \geqq 1 - \epsilon$, hence there is an $x_1$ such that $P(A \mid x_1) \geqq 1 - \epsilon$. Let $B_1 = A_{x_1}$. (Each $B_k$ will be a cylinder set with base in $\mathfrak{B}$. The base of $B_k$ will be the decoding set for $x_k$.) At the $k$th step select $x_k$ such that $P(B_k \mid x_k) \geqq 1 - \epsilon$ where

$$B_k = \bigcup_1^k A_{x_i} - \bigcup_1^{k-1} A_{x_i}.$$

This process will terminate at some $M \geqq 1$. For every $x$

$$P\left(A - A \cap \left(\bigcup_1^M A_{x_i}\right) \middle| x\right) < 1 - \epsilon$$

otherwise we could add this $x$ to $x_1, \cdots, x_M$ contradicting the definition of $M$. Thus

$$P(A) = P\left(A \cap \left(\bigcup_1^M A_{x_i}\right)\right) + P\left(A - A \cap \left(\bigcup_1^M A_{x_i}\right)\right)$$

$$\leqq \sum_1^M P(A_{x_i}) + 1 - \epsilon.$$

Now if $(x, y) \epsilon A$ then $J(x, y) > \alpha$ so that $P(y \mid x) > P(y)e^\alpha$. For fixed $x$ sum both sides of this inequality over all $y$ such that $(x, y) \varepsilon A$. Then

$$1 \geqq P(A \mid x) \geqq P(A_x)e^\alpha.$$

Thus $P(A_x) \leqq e^{-\alpha}$ for any $x \varepsilon \mathfrak{A}$ so that $P(A) \leqq Me^{-\alpha} + 1 - \epsilon$. Since $P(A) = Ge^{-\alpha} + 1 - \epsilon$, we have $M \geqq G$. Clearly the $B_1, \cdots, B_M$ are disjoint and

$$P(B_k \mid x_k) \geqq 1 - \epsilon$$

for $k = 1, \cdots, M$ so the proof is completed.

Consider a single channel $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ and let $Q(x)$ be specified and determine $P(x, y)$, $J(x, y)$. Applying Theorem 2 to $(\mathfrak{A}^{(n)}, \mathfrak{B}^{(n)}, P(v \mid u))$ and $Q(u) = Q(x_1) \cdots Q(x_n)$ with $\alpha = n(R + EJ)/2$, $G = e^{Rn}$ we see that for any $R$ such that $0 < R < EJ$ there is an $(e^{Rn}, \epsilon_n, n)$ code for $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ with

$$\epsilon_n = e^{-(EJ-R)n/2} + P\left(\frac{1}{n} J' \leqq \frac{R + EJ}{2}\right).$$

Now

$$J'(u, v) = \log \frac{P(u, v)}{P(u)P(v)} \qquad \text{if} \quad P(u, v) > 0$$

$$= 0 \qquad \text{otherwise.}$$

Let $J''(u, v) = \sum_1^n J_i(x_i, y_i)$ where

$$J_i(x_i, y_i) = \log \frac{P(x_i, y_i)}{P(x_1)P(y_1)} \qquad \text{if } P(x_i, y_i) > 0$$

$$= 0 \qquad \text{otherwise.}$$

Clearly $P(J' = J'') = 1$ and $J''$ is the sum of $n$ independent random variables each having the distribution of $J(x, y)$. Since $EJ > (R + EJ)/2$ we see that $\epsilon_n \to 0$. Now it is easily seen (and we will shortly prove even more) that for a fixed channel $EJ$ is a continuous function of $(Q(x_1), Q(x_2), \cdots, Q(x_n))$ and since the domain of the function is a closed bounded subset of Euclidean space the supremum is actually achieved. Thus for any channel $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ there is a distribution $Q(x)$ on $\mathfrak{A}$ such that $C = EJ$. Using this $Q(x)$ in the earlier portions of this paragraph we obtain the direct half of Shannon's theorem for a memoryless channel: $T \geqq C$.

By introducing a brief epsilon argument in the proof of the direct half of Shannon's theorem we could clearly have ignored the question of whether or not there is a maximizing $Q(x)$. Although the fact that there is a maximizing $Q(x)$ in the general case of a class of channels is not vital in the following work, we will pause to prove this fact now. The proof is based on Lemma 1 which will be needed later.

Lemma 1: *Let $Q(x)$, $Q'(x)$ be any two distributions on $\mathfrak{A}$ such that*

$$| Q(x) - Q'(x) | \leqq \epsilon \leqq 1/e \text{ for all } x \in \mathfrak{A}.$$

*Then*

$$| H(X) - H'(X) | \leqq a\epsilon^{1/2}$$

*where* $H(X) = -\sum_x Q(x) \log Q(x)$ *and* $H'(X) = -\sum_x Q'(x) \log Q'(x)$.

Proof: Let

$$f(y) = [-(y + \epsilon) \log (y + \epsilon)] - [-y \log y]$$

where $0 < \epsilon \leqq 1/e$ and $0 \leqq y \leqq 1 - \epsilon$. Then $f(0) = -\epsilon \log \epsilon > 0$ and $f(1 - \epsilon) = (1 - \epsilon) \log (1 = \epsilon) < 0$ also

$$f'(y) = -\log (y + \epsilon) - 1 + \log y + 1 = \log \frac{y}{y + \epsilon} < 0$$

so that $| f(y) | \leqq \max \{-\epsilon \log \epsilon, -(1 - \epsilon) \log(1 - \epsilon)\}$. Now

$$(1 - \epsilon) \log \frac{1}{1 - \epsilon} \leqq (1 - \epsilon) \left(\frac{1}{1 - \epsilon} - 1\right) = \epsilon \leqq -\epsilon \log \epsilon$$

since $\epsilon \leqq 1/e$. Thus

$$|f(y)| \leqq -\epsilon \log \epsilon = \frac{\epsilon^{\frac{1}{2}\log\frac{1}{\epsilon}}}{\left(\frac{1}{\epsilon}\right)^{\frac{1}{2}}} \leqq \epsilon^{\frac{1}{2}}$$

since $x^{1/2} - \log x \geqq 2 - \log 4 > 0$ for $x > 0$. Applying the result $|f(y)| \leqq \epsilon^{1/2}$ to $y = p$, $\epsilon = q - p$ where $0 \leqq p \leqq q \leqq 1$ and $|q - p| \leqq 1/e$ we see that

$$|[-p \log p] - [-q \log q]| \leqq (|p - q|)^{1/2}$$

which easily gives us the bound on $|H(X) - H'(X)|$ completing the proof.

LEMMA 2: *For any class of channels* $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y \mid x))$ *for* $\gamma \varepsilon \mathfrak{C}$,

$$C = \max_{Q(x)} \inf_{\gamma \varepsilon \mathfrak{C}} E_\gamma J_\gamma .$$

PROOF: Let $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ be a channel and $Q(x)$ a distribution on $\mathfrak{A}$ determining $P(x, y) = P(y \mid x)Q(x)$ and $J(x, y)$. Clearly $EJ = H(X) + H(Y) - H(X, Y)$ where $H(X) = -\sum_x P(x) \log P(x)$, $H(Y) = -\sum_y P(y) \log P(y)$, $H(X, Y) = -\sum_{x,y} P(x, y) \log P(x, y)$. Let $Q'(x)$ be another distribution on $\mathfrak{A}$ determining $P'(x, y) = P(y \mid x)Q'(x)$ and $J'(x, y)$, and note that $E'J' = H'(X) + H'(Y) - H'(X, Y)$ where the primed quantities have analogous definitions. Assume that $|Q(x) - Q'(x)| \leqq \epsilon \leqq 1/e$ for all $x \varepsilon \mathfrak{A}$. Clearly $|P(x, y) - P'(x, y)| \leqq P(y \mid x) \mid Q(x) - Q'(x)| \leqq \epsilon$ and $|P(y) - P'(y)| \leqq \sum_x |P(x, y) - P'(x, y)| \leqq a\epsilon$. Applying Lemma 1 we get

$$\begin{aligned} |EJ - E'J'| &\leqq |H(X) - H'(X)| + |H(Y) - H'(Y)| \\ &\qquad + |H(X, Y) - H'(X, Y)| \\ &\leqq a\epsilon^{1/2} + b(a\epsilon)^{1/2} + ab\epsilon^{1/2} \leqq (a + 2ab)\epsilon^{1/2}. \end{aligned}$$

Thus not only is $EJ$ continuous in $Q(x)$ but it is continuous in $Q(x)$ uniformly in $Q(x)$ and $P(y \mid x)$. We easily take $\inf_{\gamma \varepsilon \mathfrak{C}}$ on the inequalities

$$E'_\gamma J'_\gamma - (a + 2\,ab)\epsilon^{1/2} \leqq E_\gamma J_\gamma \leqq E'_\gamma J'_\gamma + (a + 2ab)\epsilon^{1/2}$$

and see that $\inf_{\gamma \varepsilon \mathfrak{C}} E_\gamma J_\gamma$ is continuous in $Q(x)$ so that once again there is a maximizing $Q(x)$ and Lemma 2 is proved.

## 4. The error bound for one channel.

THEOREM 3: *Let* $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ *be any channel. For any integer* $n$ *and any* $R > 0$ *such that* $0 \leqq C - R \leqq 1/2$, *there is an* $(e^{Rn}, \epsilon_n, n)$ *code for the channel with*

$$\epsilon_n = 2e^{-\frac{(C-R)^2}{16ab} n}.$$

PROOF: Applying Theorem 2 to $(\mathfrak{A}^{(n)}, \mathfrak{B}^{(n)}, P(v \mid u))$ with $Q(u) = Q(x_1) \cdots Q(x_n)$, where $Q(x)$ is any distribution on $\mathfrak{A}$, $G = e^{Rn}$, $\alpha = (R + \theta)n$ we see that for any $R > 0$, $\theta > 0$ there is an $(e^{Rn}, \epsilon_n, n)$ code for $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ with

$$\epsilon_n = e^{-n\theta} + P(J'' \leqq n(R + \theta))$$

where, as shown in Section 3, $J''$ is the sum of $n$ independent random variables, each having the distribution of $J(x, y)$. Select $R > 0$, $0 \leqq EJ - R \leqq 1/2$ and let $\theta = (EJ - R)^2$. Then $R + \theta \leqq R + (EJ - R)/2 = (EJ + R)/2$.

Thus it remains only to show that

$$P(J'' \leqq n(EJ + R)\tfrac{1}{2}) \leqq e^{-\frac{(EJ-R)^2}{16ab}n}$$

(we will need this result later) for we can then choose $Q$ so that $C = EJ$.

A method due to Chernoff [2] will be used to bound the probability in question. Let $0 \leqq t \leqq 1$, then

$$P\left(0 \leqq \frac{n(EJ + R)}{2} - J''\right) \leqq Ee^{t\left[\frac{n(EJ+R)}{2} - J''\right]} = e^{\frac{tn(EJ+R)}{2}} Ee^{-J''}$$
$$= [e^{\frac{t(EJ+R)}{2}} Ee^{-tJ}]^n$$

so that we need show only that for a proper selection of $t$,

$$e^{\frac{t(EJ+R)}{2}} Ee^{-tJ} \leqq e^{-\frac{(EJ-R)^2}{16ab}}.$$

Now

$$Ee^{-tJ} = 1 - tEJ + \frac{t^2}{2} EJ^2 e^{-\theta tJ}, \qquad 0 < \theta < 1.$$

We need consider only $(x, y)$ with $P(x, y) > 0$. Terms in $EJ^2 e^{-\theta tJ}$ are of the form

$$P(x,y)\left(\frac{P(x)P(y)}{P(x,y)}\right)^{\theta t} \log^2 \frac{P(x,y)}{P(x)P(y)} \leqq P(x,y)\left(\frac{1}{P(x,y)}\right)^{\theta t} \log^2 \frac{P(x,y)}{P(x)P(y)}$$
$$\leqq (P(x,y))^{1-t} \log^2 \frac{P(x,y)}{P(x)P(y)} \leqq (P(x,y))^{1-t} \log^2 P(x,y)$$

where the last inequality followed from $P(x, y) \leqq P(x)P(y)/P(x, y) \leqq 1/P(x, y)$. Also

$$[(P(x,y))^{\frac{1-t}{2}} \log P(x,y)]^2 = \left(\frac{2}{1-t}\right)^2 [(P(x,y))^{\frac{1-t}{2}} \log P(x,y))^{\frac{1-t}{2}}]^2$$
$$\leqq \left(\frac{2}{1-t}\right)^2 \frac{1}{e^2} \leqq \frac{1}{(1-t)^2}.$$

Thus

$$Ee^{-tJ} \leqq 1 - tEJ + \frac{t^2}{2}\frac{ab}{(1-t)^2} \leqq e^{-tEJ + \frac{t^2}{2}\frac{ab}{(1-t)^2}}$$

so that

$$e^{\frac{t(EJ+R)}{2}} Ee^{-tJ} \leqq e^{-\frac{1}{2}f(t)}$$

where

$$f(t) = (EJ - R)t - t^2 \frac{ab}{(1-t)^2}.$$

Let $t = (EJ - R)/4ab \leqq 1/8$ so that $1/(1-t)^2 \leqq (8/7)^2$, then

$$f\left(\frac{EJ - R}{4ab}\right) \geqq \frac{(EJ-R)^2}{4ab}\left[1 - \left(\frac{8}{7}\right)^2 \frac{1}{4}\right] \geqq \frac{(EJ-R)^2}{8ab}$$

completing the proof.

**5. The error bound for a finite set of channels.** Lemma 3 is needed in the proof of Theorem 4.

LEMMA 3: *Let* $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y \mid x))$ *for* $\gamma \in \mathfrak{C} = \{1, 2, \cdots, L\}$ *be a finite class of channels and let* $Q(x)$ *be a distribution on* $\mathfrak{A}$, *determining* $P_\gamma(x, y)$, $J_\gamma(x, y)$.

(a) *Define a channel* $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ *by* $P(y \mid x) = (1/L)\sum_{\gamma=1}^{L} P_\gamma(y \mid x)$ *and let* $Q(x)$ *determine* $P(x, y)$, $J(x, y)$. *Then for all* $\alpha$, $\delta$

$$P(J \leqq \alpha) \leqq \frac{1}{L}\sum_{\gamma=1}^{L} P_\gamma(J_\gamma \leqq \alpha + \delta) + Le^{-\delta}.$$

(b) *For any* $\alpha > 0$, $G \geqq 1$, $\delta > 0$ *there is a* $(G, \epsilon, 1)$ *code for* $\mathfrak{C}$ *with*

$$\epsilon = LGe^{-\alpha} + L^2e^{-\delta} + \sum_{1}^{L} P_\gamma(J_\gamma \leqq \alpha + \delta).$$

PROOF: We first prove part (a).

$$P(J \leqq \alpha) = \frac{1}{L}\sum P_\gamma(J \leqq \alpha) \leqq \frac{1}{L}\sum [P_\gamma(J_\gamma \leqq \alpha + \delta) + P_\gamma(J_\gamma > \alpha + \delta; J \leqq \alpha)]$$

so that we need only prove that $P_\gamma(A_\gamma) \leqq Le^{-\delta}$ where $A_\gamma = (J_\gamma \alpha + \delta; J \leqq \alpha)$. For any $(x, y) \in A_\gamma$ with $P_\gamma(x, y) > 0$ we have

$$e^{\alpha}P(y) \geqq P(y \mid x) \geqq \frac{1}{L} P_\gamma(y \mid x) \geqq \frac{1}{L} e^{\alpha+\delta} P_\gamma(y)$$

so that $P_\gamma(y) \leqq Le^{-\delta}P(y)$. Summing this last inequality over all $y$ such that there is an $x$ with $(x, y) \in A_\gamma$ we get $P_\gamma(A_\gamma) \leqq \sum P_\gamma(y) \leqq Le^{-\delta}$ which completes the proof of part (a).

Applying Theorem 2 to the channel $P(y \mid x)$ defined in part (a) and then using part (a) to bound $P(J \leqq \alpha)$ we find that there is a $(G, \epsilon_0, 1)$ code for $P(y \mid x)$ with

$$\epsilon_0 = Ge^{-\alpha} + P(J \leqq \alpha) \leqq Ge^{-\alpha} + \frac{1}{L}\sum_{\gamma} P_\gamma(J_\gamma \leqq \alpha + \delta) + Le^{-\delta}.$$

Now $P_\gamma(y \mid x) \leqq LP(y \mid x)$ so that if $x_i$ is an input letter for the $(G, \epsilon_0, 1)$ code and $B_i$ is its decoding set, then $P_\gamma(B_i^c \mid x_i) \leqq L\, P_\gamma(B_i^c \mid x_i) \leqq L\epsilon_0$. Thus the $(G, \epsilon_0, 1)$ code for $P(y \mid x)$ is a $(G, L\epsilon_0, 1)$ code for $\mathfrak{C}$ and the lemma is proved.

THEOREM 4: *Let* $(\mathfrak{A}, \mathfrak{B}, P_\gamma(y \mid x))$ *for* $\gamma \in \mathfrak{C} = \{1, 2, \cdots, L\}$ *be a finite class of channels. For any* $R > 0$ *such that* $0 \leqq C - R \leqq 1/2$ *there is an* $(e^{Rn}, \epsilon_n, n)$ *code with*

$$\epsilon = 2L^2 e^{-\frac{(C-R)^2}{16ab}n}.$$

PROOF. Applying part (b) of Lemma 3 to the class of channels $(\mathfrak{A}^{(n)}, \mathfrak{B}^{(n)}, P_\gamma(v \mid u))$ with $Q(u) = Q(x_1) \cdots Q(x_n)$ and $Q(x)$ a distribution for which $C = \inf_{\gamma \in \mathfrak{C}} E_\gamma J_\gamma$ and $G = e^{Rn}$, $\alpha = (R + \theta/2)n$, $\delta = \theta n/2$ we see that there is an $(e^{Rn}, \epsilon_n, n)$ code for $\mathfrak{C}$ with

$$\epsilon_n = (L + L^2)e^{-\frac{\theta}{2}n} + \sum_1^L P_\gamma\left(\frac{1}{n} J_\gamma \leqq R + \theta\right).$$

Let $\theta = (C - R)^2$ and note that $R + (C - R)^2 \leqq R + (C - R)/2 \leqq R + (E_\gamma J_\gamma - R)/2 = (E_\gamma J_\gamma + R)/2$. Thus,

$$\epsilon_n \leqq (L + L^2)e^{-\frac{(C-R)^2}{16ab}n} + \sum_1^L P_\gamma\left(\frac{1}{n} J_\gamma \leqq \tfrac{1}{2}(R + E_\gamma J_\gamma)\right).$$

Now

$$P_\gamma\left(\frac{1}{n} J_\gamma \leqq \tfrac{1}{2}(R + E_\gamma J_\gamma)\right) \leqq P_\gamma\left(\frac{1}{n} J_\gamma \leqq \tfrac{1}{2}(R' + E_\gamma J_\gamma)\right)$$

where $R' = E_\gamma J_\gamma - (C - R) \geqq R$ and $0 \leqq E_\gamma J_\gamma - R' \leqq 1/2$. Therefore, we can apply the result obtained in the proof of Theorem 3 and get

$$P_\gamma\left(\frac{1}{n} J_\gamma \leqq \tfrac{1}{2}(R' + E_\gamma J_\gamma)\right) \leqq e^{-\frac{(E_\gamma J_\gamma - R')^2}{16ab}n} = e^{-\frac{(C-R)^2}{16ab}n}.$$

Now $L \geqq 2$ so that $2L + L^2 = L(L + 2) \leqq 2L^2$ and since Theorem 4 reduces to Theorem 3 for $L = 1$, the proof is completed.

**6. The direct half of Theorem 1.** Lemmas 4 and 5 are needed for the proof of part (a) of Theorem 1.

LEMMA 4: *Let* $\mathfrak{A}$, $\mathfrak{B}$ *be given. For every integer* $M \geqq 2b^2$ *there is a class of channels* $(\mathfrak{A}, \mathfrak{B}, P_j(y \mid x))$ *with* $\in \mathfrak{D}_M$, *where* $\mathfrak{D}_M$ *has at most* $(M + 1)^{ab}$ *elements, such that for any channel* $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ *there is a channel* $(\mathfrak{A}, \mathfrak{B}, P'(y \mid x))$ *in* $\mathfrak{D}_M$ *such that*:

(a) $|P(y \mid x) - P'(y \mid x)| \leqq b/M$ *for all* $x, y$.

(b) $P(y \mid x) \leqq e^{2b^2/M} P'(y \mid x)$ *for all* $x, y$.

(c) *For any distribution* $Q(x)$ *on* $\mathfrak{A}$ *let* $P(x, y) = P(y \mid x)Q(x)$, $P'(x, y) = P'(y \mid x)Q(x)$, *then*

$$|EJ - E'J'| \leqq 2b\left(\frac{b}{M}\right)^{1/2}.$$

PROOF. Let $\mathfrak{D}_M$ be the class of channels $(\mathfrak{A}, \mathfrak{B}, P(y \mid x))$ such that for all $x, y$ we have $MP(y \mid x) =$ an integer. Clearly $\mathfrak{D}_M$ has at most $(M + 1)^{ab}$ elements. Given the distributions $P(y \mid x)$ we will first construct $P'(y \mid x)$ and prove (a),

(b). For this purpose it is enough to carry out the construction for one $x_0$. Arrange the "b" numbers $P(y \mid x_0)$ in ascending order and designate them by $p_1 \leqq p_2 \leqq \cdots \leqq p_b$. For $i = 1, \cdots, (b-1)$ select $p_i'$ uniquely by $p_i \leqq p_i' < p_i + 1/M$, $Mp_i' =$ an integer. $p_i'$ will be $P'(y \mid x_0)$ with the $y$ being the one corresponding to $p_i$. Clearly

$$p_i \leqq e^{\frac{2b^2}{M}} p_i' \quad \text{and} \quad |p_i - p_i'| \leqq \frac{b}{M}$$

for $i = 1, \cdots, (b-1)$. It remains to show that if $p_b' = 1 - \sum_1^{b-1} p_i'$ then $p_b' \geqq 0$ and $p_b'$, $p_b$ satisfy the same relations. Now

$$p_b' \geqq 1 - \sum_1^{b-1} \left(p_i + \frac{1}{M}\right) \geqq p_b - \frac{b}{M} \geqq \frac{1}{b} - \frac{b}{M} \geqq \frac{1}{b} - \frac{1}{2b} = \frac{1}{2b}.$$

Thus $p_1', \cdots, p_b'$ form a distribution and $p_b \geqq p_b' \geqq p_b - b/M$ so that

$$|p_b - p_b'| \leqq b/M.$$

Also

$$p_b \leqq p_b' + \frac{b}{M} \leqq p_b' + \frac{2b^2}{M}\frac{1}{2b} \leqq p_b'\left(1 + \frac{2b^2}{M}\right) \leqq e^{\frac{2b^2}{M}} p_b'$$

completing the proof of parts (a) and (b).

In the proof of part (c) we will use part (a) and Lemma 1. In order to use Lemma 1 we observe that $b/M \leqq 1/2b \leqq 1/4 < 1/e$. We also note that

$$|P(y) - P'(y)| \leqq \sum_x |P(y \mid x) - P'(y \mid x)| Q(x) \leqq b/M.$$

Now

$$\begin{aligned} |EJ - E'J'| &\leqq \left| \left[-\sum_y P(y) \log P(y)\right] - \left[-\sum_y P'(y) \log P'(y)\right] \right| \\ &+ \left| \left[-\sum_{x,y} P(x,y) \log P(x,y)\right] - \left[-\sum_{x,y} P'(x,y) \log P'(x,y)\right] \right| \leqq b\left(\frac{b}{M}\right)^{1/2} \\ &+ \sum_x Q(x) \left| \left[-\sum_y P(y \mid x) \log P(y \mid x)\right] - \left[-\sum_y P'(y \mid x) \log P'(y \mid x)\right] \right| \\ &\leqq b\left(\frac{b}{M}\right)^{1/2} + b\left(\frac{b}{M}\right)^{1/2} \end{aligned}$$

and the lemma is proved.

LEMMA 5: *Let $(\mathcal{A}, \mathcal{B}, P'(y \mid x))$, $(\mathcal{A}, \mathcal{B}, P(y \mid x))$ be two channels and $A$ a non-negative number such that $P(y \mid x) \leqq e^A P'(y \mid x)$ for all $x, y$. Any $(e^{Rn}, \epsilon_n, n)$ code for $(\mathcal{A}, \mathcal{B}, P'(y \mid x))$ is an $(e^{Rn}, \epsilon_n e^{An}, n)$ code for $(\mathcal{A}, \mathcal{B}, P(y \mid x))$.*

PROOF: Let $u = (x_1, \cdots, x_n) \in \mathcal{A}^{(n)}$, $v = (y_1, \cdots, y_n) \in \mathcal{B}^{(n)}$. Then

$$P(v \mid u) = \prod_1^n P(y_i \mid x_i) \leqq e^{An} \prod_1^n P'(y_i \mid x_i) = e^{An} P'(v \mid u).$$

Thus for any subset $D$ of $\mathcal{B}^{(n)}$ and any $u \in \mathcal{A}^{(n)}$ we have

$$P(D \mid u) \leqq e^{An} P'(D \mid u).$$

Let $u_i \varepsilon \mathfrak{A}^{(n)}$ be an input message and $B_i$ the corresponding decoding set of an $(e^{Rn}, \epsilon_n, n)$ code for $(\mathfrak{A}, \mathfrak{B}, P'(y \mid x))$. Then

$$P(B_i^c \mid u_i) \leqq e^{An} P'(B_i^c \mid u_i) \leqq e^{An} \epsilon_n$$

and the proof is completed.

We turn now to the proof of part (a) of Theorem 1. For each $P(y \mid x) \varepsilon \mathfrak{C}$ select a $P'(y \mid x) \varepsilon \mathfrak{D}_M$ according to Lemma 4 and let $\mathfrak{C}'$ denote this set of channels. Let $C' = C(\mathfrak{C}')$. Since $\mathfrak{C}'$ has at most $(M + 1)^{ab}$ elements we know from Theorem 4 that if $R' > 0$, $0 \leqq C' - R' \leqq 1/2$ then there is an $(e^{R'n}, \epsilon_n, n)$ code for $\mathfrak{C}'$ with

$$\epsilon_n' = 2(M + 1)^{2ab} e^{-\frac{(C'-R')^2}{16ab}n}.$$

For each $P(y \mid x) \varepsilon \mathfrak{C}$ there is a $P'(y \mid x) \varepsilon \mathfrak{C}'$ such that

$$P(y \mid x) \leqq e^{\frac{2b^2}{M}} P'(y \mid x)$$

so that from Lemma 5 the code which we have for $\mathfrak{C}'$ is an $(e^{R'n}, \epsilon_n, n)$ code for $\mathfrak{C}$ with

$$\epsilon_n = 2(M + 1)^{2ab} \exp -\left\{\frac{(C' - R')^2}{16ab} - \frac{2b^2}{M}\right\} n.$$

Let $C = C(\mathfrak{C})$ and let $Q(x)$ be a maximizing distribution for $\mathfrak{C}$. We wish to show that $C'$ cannot be very much smaller than $C$. For every $P'(y \mid x) \varepsilon \mathfrak{C}'$ there is a $P(y \mid x) \varepsilon \mathfrak{C}$ such that $EJ \leqq E'J' + 2b(b/M)^{1/2}$ where we use $Q(x)$ in both cases. Thus for every $P'(y \mid x) \varepsilon \mathfrak{C}$

$$C = \inf_{\mathfrak{C}} EJ \leqq E'J' + 2b\left(\frac{b}{M}\right)^{1/2}$$

so that

$$C \leqq \inf_{\mathfrak{C}'} E'J' + 2b\left(\frac{b}{M}\right)^{1/2} \leqq C' + 2b\left(\frac{b}{M}\right)^{1/2}.$$

Let $R > 0$ be given such that $0 < C - R \leqq 1/2$. We must show how to select $R'$ and $M$ to get our result into the final form.

We select an integer $M$ such that

$$\frac{2^8ab^3}{(C - R)^2} \leqq M \quad \text{and} \quad (M + 1) \leqq \frac{2^9ab^3}{(C - R)^2}$$

so that

$$2b\left(\frac{b}{M}\right)^{1/2} \leqq \frac{C - R}{2} \quad \text{and} \quad \frac{2b^2}{M} \leqq \frac{(C - R)^2}{2^7ab}.$$

We define $R'$ by

$$C' - R' = C - R - 2b\left(\frac{b}{M}\right)^{1/2} \geqq \frac{C - R}{2} > 0.$$

Clearly $C' - R' \leq 1/2$ so that we have an $(e^{R'n}, \epsilon_n, n)$ code for $\mathcal{C}$ with

$$\epsilon_n \leq 2(M+1)^{2ab} \exp -\left\{\frac{(C-R)^2}{4(16ab)} - \frac{(C-R)^2}{2^7 ab}\right\}$$

$$\leq 2\left[\frac{2^9 ab^3}{(C-R)^2}\right]^{2ab} \exp -\left\{\frac{(C-R)^2}{2^7 ab}\right\}.$$

The inequality $C \leq C' + 2b(b/M)^{1/2}$ shows that $R' \geq R$ and an $(e^{R'n}, \epsilon_n, n)$ code for $\mathcal{C}$ can easily be reduced to an $(e^{Rn}, \epsilon_n, n)$ code for $\mathcal{C}$ so that part (a) of Theorem 1 is proved.

**7. Converse half of Theorem 1.** The proof is based on Lemma 6.

LEMMA 6: *Let $G$ be an integer, $\mathcal{A}$ a finite set and let $u_1, \cdots, u_G$ be distinct elements of $\mathcal{A}^{(n)}$. Define $Q(x)$ on $\mathcal{A}$ by*

$$Q(x) = \frac{1}{nG}\sum_{j=1}^{G} (\text{the number of times that } x \text{ appears in } u_j).$$

*Then any $(G, \epsilon, n)$ code, for a channel $(\mathcal{A}, \mathcal{B}, P(y \mid x))$ which uses these $u_1, \cdots, u_G$ for inputs must satisfy*

$$(1-\epsilon)\log G - \log 2 \leq nEJ$$

*where $Q(x)$ is used to define $P(x, y)$ and $J(x, y)$.*

PROOF: Define a distribution $\nu(u)$ on $\mathcal{A}^{(n)}$ by $\nu(u) = 1/G$ if $u$ is one of $u_1, \cdots, u_G$ and $\nu(u) = 0$ otherwise. Define a distribution $P(u, v)$ on $\mathcal{A}^{(n)} \times \mathcal{B}^{(n)}$ by $P(u, v) = P(v \mid u)\nu(u)$ where $P(v \mid u)$ is obtained from the $n$-extension of $(\mathcal{A}, \mathcal{B}, P(y \mid x))$. Now define $n$ distributions on $\mathcal{A} \times \mathcal{B}$ by

$$P^{(i)}(x, y) = P(y \mid x)\nu^{(i)}(x)$$

for $i = 1, \cdots, n$ where

$$\nu^{(i)}(x) = \sum_{x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n} \nu(x_1, \cdots, x_{i-1}, x, x_{i+1}, \cdots, x_n)$$

and observe that $Q(x) = (1/n)\sum_{i=1}^{n} \nu^{(i)}(x)$. Thus, the lemma will be proved if the following chain of inequalities is proved.

$$(1-\epsilon)\log G - \log 2 \leq \sum_{u,v} P(u, v)\log\frac{P(u, v)}{P(u)P(v)}$$

$$\leq \sum_{i=1}^{n}\sum_{x,y} P^{(i)}(x, y)\log\frac{P^{(i)}(x, y)}{P^{(i)}(x)P^{(i)}(y)} \leq n\sum_{x,y} P(x, y)\log\frac{P(x, y)}{P(x)P(y)}.$$

Using $\log x = (\log 2)\log_2 x$ to convert a result from Feinstein [8], pp. 29, 39, 44; which is due to Fano [5]; we obtain the first inequality. The second inequality follows from page 30 of Feinstein [8]. We proceed to prove the third inequality. Now

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{x,y} P^{(i)}(x, y)\,[\log P(y \mid x) - \log P^{(i)}(y)]$$

$$= \sum_{x,y} P(x, y)\log P(y \mid x) - \frac{1}{n}\sum_{i=1}^{n}\sum_{y} P^{(i)}(y)\log P^{(i)}(y)$$

but

$$-\frac{1}{n}\sum_{i=1}^{n}\sum_{y} P^{(i)}(y)\log P^{(i)}(y) \leqq -\sum_{y}\left(\frac{1}{n}\sum_{i=1}^{n} P^{(i)}(y)\right)\log\left(\frac{1}{n}\sum_{i=1}^{n} P^{(i)}(y)\right)$$
$$= -\sum_{y} P(y)\log P(y) = -\sum_{x,y} P(x,y)\log P(y)$$

where this last inequality follows from Lemma 4 on page 16 of Feinstein [8]. Combining the above, we complete the proof of the third inequality and hence of the lemma.

From Lemma 6 we immediately obtain that if $G$ is an integer then for any $(G, \epsilon, n)$ code for a class $\mathfrak{C}$ of channels there is a $Q(x)$ on $\mathfrak{A}$ such that

$$(1-\epsilon)\log G - \log 2 \leqq n \inf_{\gamma\epsilon\mathfrak{C}} E_\gamma J_\gamma \leqq nC.$$

Now $e^{Rn}$ may not be an integer but

$$\log\,[e^{Rn}] \geqq \log\,(e^{Rn}-1) \geqq nR + \log(1-e^{-Rn}) \geqq nR - \log 2$$

so that

$$(1-\epsilon)\,(nR - \log 2) \leqq nC + \log 2$$

which completes the proof of part (b) of Theorem 1.

## REFERENCES


[1] David Blackwell, Leo Breiman, A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1209–1220.

[2] Herman Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[3] Peter Elias, "Coding for noisy channels," *I.R.E. Convention Record* (1955), part 4, pp. 37–44.

[4] Peter Elias, "Coding for two noisy channels," *Proceedings of the London Symposium on Information Theory*, Butterworth Scientific Publications, London, 1955.

[5] R. M. Fano, "Statistical theory of communication," notes on a course given at the Massachusetts Institute of Technology, 1952, 1954.

[6] Amiel Feinstein, "A new basic theorem of information theory," *I.R.E. Trans. P.G.I.T.*, September, 1954, pp. 2–22.

[7] Amiel Feinstein, "Error bounds in noisy channels without memory," *I.R.E. Trans. P.G.I.T.*, September, 1955, pp. 13–14.

[8] Amiel Feinstein, *Foundations of Information Theory*, McGraw-Hill, New York, 1958.

[9] A. I. Khinchin, "On the fundamental theorems of information theory," *Uspekhi Mathematicheskikh Nauk.*, Vol. 21 (1956), pp. 17–75.

[10] A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications, Inc., 1957.

[11] Brockway McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, Vol. 24(1953), pp. 196–219.

[12] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379–423, and 623–656.

[13] Claude E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1957), pp. 6–25.

[14] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois J. Math.*, Vol. 1 (1957), pp. 591–606.

# INFINITE CODES FOR MEMORYLESS CHANNELS

By David Blackwell[1]

*University of California, Berkeley*

**1. Introduction and summary.** For a memoryless channel with finite input alphabet $A$, finite output alphabet $B$, and probability law $p(b \mid a)$, the *capacity* $C$ is defined as the maximum over all probability distributions $q$ on $A$ of

$$\sum_{a,b} q(a)p(b \mid a)\log_2(p(b \mid a)/\sum_a q(a)p(b \mid a)).$$

Shannon [1] has obtained the following result.

*Exponential error bound. For any $C_0 < C$ there is a number $\rho < 1$ such that, for every positive integer $N$, there is a set $S \subset A^{(N)}$ with at least $2^{C_0 N}$ elements and a function $g$ from $B^{(N)}$ to $S$, such that, for every $s = (a_1, \cdots, a_N) \varepsilon S$,*

$$\sum p(b_1 \mid a_1) \cdots p(b_N \mid a_N) < 2\rho^N,$$

where the sum extends over all sequences $b_1, \cdots, b_N$ for which $g(b_1, \cdots, b_N) \neq s$.

Thus if the sender selects any $s \varepsilon S$ and places its letters $a_1, \cdots, a_N$ successively into the channel, and the receiver, on observing the resulting output sequence $b_1, \cdots, b_N$, decides that the input was $g(b_1, \cdots, b_N)$, the probability that he makes an error is less than $2\rho^N$, no matter what $s \varepsilon S$ was chosen. This result may be described as follows: it is possible to transmit at any rate $C_0 < C$, with arbitrarily small probability of error, by using block codes of sufficient length.

We wish to draw a slightly stronger conclusion, as follows. We imagine an infinite sequence $x = (x_1, x_2, \cdots)$ of 0's and 1's, which we are required to transmit across the channel. At time $N$, the sender will have observed the first $[C_0 N]$ coordinates of $x$, and will place the $N$th input symbol in the channel. The receiver, having at this point observed the first $N$ channel outputs, will estimate the first $M(N)$ coordinates of $x$. If $M(N)/C_0 N \to 1$ as $N \to \infty$ and if, for every $x$, all but a finite number of his estimates are correct (i.e., agree with $x$ in every coordinate estimated) with probability 1, we shall say that the channel is being used at rate $C_0$. Our result is that, in this sense, a (memoryless) channel can be used at any rate $C_0 < C$.

The result stated below is exactly this result, for the special case $C_0 = 1$. The general case involves no new ideas, but only more notation, and we shall restrict attention to the case $C_0 = 1$. The function $f_n$ of a code, as defined below, specifies the $n$th channel input symbol, as a function of the first $n$ coordinates of $x$. The number $M(n)$ is the number of $x$ coordinates to be estimated by the

Received March 27, 1959; revised June 1, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53). This paper in whole or in part may be reproduced for any purpose of the United States Government.

receiver after observing the first $n$ output symbols, and the function $g_n$ specifies the estimate.

We now state the result precisely.

For any finite set $S$, we denote by $S^{(N)}$ the set of all sequences $(s_1, \cdots, s_N)$, where $s_n \varepsilon S$ for $n = 1, 2, \cdots, N$. For a memoryless channel with finite input alphabet $A$, finite output alphabet $B$, an infinite code (for transmitting at rate 1) is defined as consisting of (a) a sequence $\{f_n\}$ of functions, where $f_n$ maps $I^{(n)}$ into $A$, and $I$ consists of the two elements 0 and 1, (b) a nondecreasing sequence $\{M(n)\}$ of positive integers such that $M(n)/n \to 1$ as $n \to \infty$, and (c) a sequence $\{g_n\}$ of functions, where $\{g_n\}$ maps $B^{(n)}$ into $I^{(M(n))}$.

An infinite sequence $x = (x_1, x_2, \cdots)$ of 0's and 1's, together with an infinite code, defines a sequence of independent output variables $y_1, y_2, \cdots$, with

$$\Pr\{y_n = b\} = p(b \mid f_n(x_1, \cdots, x_n)),$$

where $p(b \mid a)$ is the probability that the output symbol of the channel is $b$, given that the corresponding input symbol is $a$, and defines a sequence of estimated messages $t_1, t_2, \cdots$, where $t_n = g_n(y_1, \cdots, y_n)$. We shall say that the code is effective at $x$ if, with probability 1,

$$t_n = (x_1, \cdots, x_{M(n)})$$

for all sufficiently large $n$, and shall say that the code is effective if it is effective for every $x$. The result of this note is the

THEOREM: *For any memoryless channel with capacity $C > 1$, there is an effective code.*

**2. Proof of the theorem.** Choose a number $D$ with $1 < D < C$, and let $\rho$ be the number $<1$ which Shannon's exponential error bound associates with transmitting at rate $D$. Thus we can, for any positive integer $R$, transmit any $[DR]$ $x$-coordinates with $R$ uses of the channel, with error probability at most $2\rho^R$. We shall divide the $x$-sequence into successive blocks, of length $R(1), R(2), \cdots$, where $\{R(k)\}$ is an appropriately chosen increasing sequence of positive integers. We may use the channel, during the time the $k + 1$st block of $x$-symbols is observed, to transmit up to $[DR(k + 1)]$ $x$-coordinates, among those received to date, with error probability at most $2\rho^{R(k+1)}$. We choose to transmit the $k$th block, containing $R(k)$ $x$-coordinates, and to repeat the first $Q(k)$ coordinates of $x$, where $\{Q(k)\}$ is a nondecreasing sequence of nonnegative integers such that

$$Q(k) + R(k) \leqq [DR(k + 1)],$$
$$Q(k) \leqq R(1) + \cdots + R(k - 1).$$

Since $\{R(k)\}$ is strictly increasing, $\sum_k \rho^{R(k)}$ converges, so that, with probability 1, only a finite number of errors will be committed. That is to say, the receiver, after observing the $k + 1$st block of output symbols, estimates the first $Q(k)$ $x$-symbols, say as $u(k)$, and the $k$th block of $x$-symbols, say as $v(k)$, and we have, with probability 1,

$$u(k) = c(k), \quad v(k) = d(k)$$

for all sufficiently large $k$, where $c(k)$ denotes the first $Q(k)$ coordinates of $x$ and $d(k)$ denotes the $k$th block of $x$-coordinates. After observing the $k + 1$st block of output symbols and making the estimates $u(k)$, $v(k)$, the receiver will have estimated each of the first $R(1) + \cdots + R(k) = T(k)$ coordinates of $x$ at least once. He now forms an estimate $w(k)$ of the first $T(k)$ coordinates, using the latest estimate made on each coordinate. If

$$Q(k) = R(1) + \cdots + R(i-1) + h, 0 \leqq h < R(i),$$

the estimate $w(k)$ is:

$$w(k) = (u(k), v^*(i), v(i+1), \cdots, v(k)),$$

where $v^*(i)$ consists of the last $R(i) - h$ coordinates of $v(i)$. If $Q(k) \to \infty$ with $k$, so does $i$. Since, with probability 1, all $u(i)$, $v(i)$ for $i$ sufficiently large are correct, we conclude that, with probability 1,

$$w(k) = (x_1, \cdots, x_{T(k)})$$

for all sufficiently large $k$. We have thus defined a sequence $\{w(k)\}$ of estimates, where $w(k)$ estimates the first $T(k)$ coordinates of $x$ after $T(k+1)$ outputs have been received, such that, with probability 1, all but a finite number of $w(k)$ are correct.

For $n < T(2)$, we define $g_n$ arbitrarily; for $T(k+1) \leqq n < T(k+2)$, we define $g_n$ as $w(k)$. Thus, for $T(k+1) \leqq n < T(k+2)$, we have $M(n) = T(k)$, and $M(n)/n \to 1$ as $n \to \infty$ if $T(k)/T(k+2) \to 1$ as $k \to \infty$.

In summary, any two sequences $\{R(k)\}$, $\{Q(k)\}$ can be used to define an effective code, if

(1) $\{R(k)\}$ is a strictly increasing sequence of positive integers.
(2) $\{Q(k)\}$ is a nondecreasing sequence of nonnegative integers.
(3) $Q(k) + R(k) \leqq [DR(k+1)]$.
(4) $Q(k) \leqq R(1) + \cdots + R(k-1)$.
(5) $Q(k) \to \infty$ as $k \to \infty$.
(6) $(R(1) + \cdots + R(k))/(R(1) + \cdots + R(k+2)) \to 1$ as $k \to \infty$.

The sequences $R(k) = k$, $Q(k) = [\min(1, D-1)(k-1)]$, for instance, satisfy (1) $\cdots$ (6).

This completes the proof.

It would be desirable to extend the theorem to finite-state channels. The method of this paper relies on Shannon's exponential error bounds, and such bounds are not yet known for general finite-state channels.

## REFERENCE

[1] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1956), pp. 6–25.

# NOTES

## A PROOF OF WALD'S THEOREM ON CUMULATIVE SUMS

By N. L. Johnson

*University College London*

**1. Introduction.** In the theory of sequential analysis developed by Wald [1], there occurs a theorem, one form of which can be expressed as follows:

THEOREM 1. *If*

(i) $z_1, z_2, z_3 \cdots$ *are independent random variables with common expected value* $\mathcal{E}(z) = \mu$,

(ii) $\mathcal{E}(|z_i|) \leqq A < \infty$ *for all* $i$, *and some finite* $A$,

(iii) $n$ *is a random variable taking values* 1, 2, 3, $\cdots$ *with probabilities* $P_1, P_2, P_3 \cdots$ *respectively, and*

(iv) *the event* $\{n \geqq i\}$ *depends only on* $z_1, z_2, \cdots z_{i-1}$,

*then, setting* $Z_n = \sum_{i=1}^{n} z_i$,

$$\mathcal{E}(Z_n) = \mu\mathcal{E}(n).$$

This note presents a simple proof of this theorem. It appears to be an abbreviated form of an argument due to Wolfowitz [2].

In Sections 3 and 4 of this note an extension of the method to the evaluation of the variance of $n$ is discussed.

**2. Proof of Theorem 1.** Let $y_i = 1$ if $z_i$ is observed (i.e. if the event $\{n \geqq i\}$ occurs) and $y_i = 0$ if $\{n \geqq i\}$ is not observed, so that

$$\Pr\{y_i = 1\} = \Pr\{n \geqq i\} = \textstyle\sum_{j=i}^{\infty} P_j .$$

Then $Z_n = \sum_{i=1}^{\infty} y_i z_i$ and $\mathcal{E}(Z_n) = \mathcal{E}(\sum_{i=1}^{\infty} y_i z_i) = \sum_{i=1}^{\infty} \mathcal{E}(y_i z_i)$ since $\sum_{i=1}^{\infty} |\mathcal{E}(y_i z_i)| < A\mathcal{E}(n) < \infty$. By reason of (iv),

$$\mathcal{E}(y_i z_i) = \mathcal{E}(y_i)\mathcal{E}(z_i),$$

so

$$\mathcal{E}(Z_n) = \sum_{i=1}^{\infty} \mathcal{E}(y_i)\mathcal{E}(z_i) = \mu \sum_{i=1}^{\infty} \mathcal{E}(y_i),$$

$$= \mu \sum_{i=1}^{\infty} (P_i + P_{i+1} + \cdots) = \mu \sum_{i=1}^{\infty} iP_i = \mu\mathcal{E}(n).$$

**3. An analogous second moment theorem.**

THEOREM 2. *If we assume, in addition to* (i)-(iv), *that*

(v) $\operatorname{var}(z_i) = \mathcal{E}(z_i^2) - \mu^2 = \sigma^2 < \infty$, *with the same value for all* $i$,

Received July 14, 1958; revised February 10, 1959.

(vi) $\mathcal{E}[(z_j - \mu)^2 \mid n \geqq i] \leqq B < \infty$ *for all* $j < i$, *with* $B$ *independent of* $i$,

(vii) $\mathcal{E}(n^2) = \sum_{i=1}^{\infty} i^2 P_i < \infty$, *then, setting* $Z'_n = Z_n - n\mu$

$$\mathcal{E}(Z'^2_n) = \sigma^2 \mathcal{E}(n).$$

Proof. Let $z'_i = z_i - \mu$, so $\mathcal{E}(z'^2_i) = \sigma^2$. Then $Z'_n = \sum_{i=1}^{n} z'_i$ and $\mathcal{E}(Z'^2_n) = \mathcal{E}(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} y_i z'_i y_j z'_j)$. Since

$$\begin{aligned}
\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathcal{E}(\,|\, y_i z'_i y_j z'_j \,|\,) &= \sum_{i=1}^{\infty} \mathcal{E}(y_i z'^2_i) + 2\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} \mathcal{E}(y_i \,|\, z'_i z'_j \,|\,) \\
&= \sigma^2\, \mathcal{E}(n) + 2\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} \mathcal{E}(y_i\, \mathcal{E}(\,|\, z'_i z'_j \,|\, \mid n \geqq i)) \\
&\leqq \sigma^2 \mathcal{E}(n) + 2\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} \mathcal{E}(y_i)[\mathcal{E}(z'^2_i)\mathcal{E}(z'^2_j \mid n \geqq i)]^{\frac{1}{2}} \\
&\leqq \sigma^2 \mathcal{E}(n)\; 2\sigma B^{\frac{1}{2}} \sum_{i=1}^{\infty}\sum_{j=1}^{i-1} \mathcal{E}(y_i) \\
&= \sigma^2 \mathcal{E}(n) + 2\sigma B^{\frac{1}{2}} \sum_{i=1}^{\infty} \frac{1}{2} i(i-1)P_i \\
&\leqq \sigma^2 \mathcal{E}(n) + \sigma B^{\frac{1}{2}}\, [\mathcal{E}(n^2) - \mathcal{E}(n)] \\
&< \infty,
\end{aligned}$$

we can invert the order of summation and expectation, giving

$$\begin{aligned}
\mathcal{E}(Z'^2_n) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathcal{E}(y_i z'_i y_j z'_j) \\
&= \sum_{i=1}^{\infty} \mathcal{E}(y_i z'^2_i) + 2\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} \mathcal{E}(y_i z'_i z'_j) \\
&= \sum_{i=1}^{\infty} \mathcal{E}(y_i)\mathcal{E}(z'^2_i) + 2\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} \mathcal{E}(y_i z'_j)\mathcal{E}(z'_i) \\
&= \sigma^2 \mathcal{E}(n),
\end{aligned}$$

since $\mathcal{E}(z'_i) = 0$, and by reason of (iv).

**4. The variance of** $n$. If, now, we make the assumption (viii) $\mathcal{E}(Z_n \mid n)$ is independent of $n$, we have, using Theorem 2,

$$\begin{aligned}
\sigma^2 \mathcal{E}(n) = \mathcal{E}(Z'^2_n) &= \mathcal{E}[(Z_n - n\mu)^2] \\
&= \mathcal{E}(Z^2_n) - 2\mathcal{E}(n)\mu\mathcal{E}(Z_n) + \mathcal{E}(n^2)\mu^2 \\
&= \mathcal{E}(Z^2_n) - 2[\mu\mathcal{E}(n)]^2 + \mathcal{E}(n^2)\mu^2
\end{aligned}$$

(using Theorem 1).

Hence

$$\mathcal{E}(n^2) = [\sigma^2\mathcal{E}(n) - \mathcal{E}(Z_n^2)]\mu^{-2} + 2[\mathcal{E}(n)]^2$$

or

$$\begin{aligned} \text{var}(n) &= [\sigma^2\mathcal{E}(n) - \mathcal{E}(Z_n^2)]\mu^{-2} + [\mathcal{E}(n)]^2 \\ &= [\sigma^2\mathcal{E}(n) - \text{var}(Z_n)]\mu^{-2} \end{aligned}$$

**5. Concluding remarks.** Theorem 2 has been stated in [3] with the weaker condition

$$\text{(vi)}' \qquad \mathcal{E}(n^{\frac{3}{2}}) < \infty$$

in place of conditions (vi) and (vii), but an error in the proof was pointed out in [4].

Conditions (vi) and (vii) may be replaced by *either*

$$\text{(vi)}'' \qquad \mathcal{E}(n^{2+\delta}) < \infty \qquad (\delta > 0)$$

*or*

$$\text{(vi)}''' \qquad \textstyle\sum_{i=1}^{\infty} \sqrt{P_i} < \infty.$$

Condition (vi) is certainly satisfied if the event $\{n \geqq i\}$ is equivalent to $a < Z_j < b$ for all $j < i$. For then we must have $|z_j - \mu| < b - a + |\mu|$ and so $\mathcal{E}[(z_j - \mu)^2 \mid n \geqq i] < (b - a + |\mu|)^2$. This condition is therefore satisfied in standard sequential procedures, which have continuation regions of form $a < Z_j < b$. Condition (vii) is also satisfied by such procedures when (v) is satisfied (see [5]).

I would like to thank the referee and Professor W. Hoeffding for help in the presentation of this paper; in particular, they suggested the alternative conditions (vi)″ and (vi)‴.

REFERENCES

[1] Abraham Wald, "On cumulative sums of random variables," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 283–96.

[2] J. Wolfowitz, "The efficiency of sequential estimates, and Wald's equation for sequential processes", *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–30.

[3] A. N. Kolmogorov and Y. Prohorov, "On sums of a random number of random terms," *Uspehi Matem. Nauk (N.S.)*, Vol. 4 (1949), pp. 168–72 (In Russian).

[4] J. Seitz and K. Winkelbauer, "Remarks concerning a paper of Kolmogorov and Prohorov," *Czechoslovak Math. J.*, Vol. 3 (1953), pp. 89–91 (In Russian, with English summary).

[5] Charles Stein, "A note on cumulative sums," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 498–9.

# A NOTE ON MULTIPLE INDEPENDENCE UNDER MULTIVARIATE NORMAL LINEAR MODELS

By V. P. Bhapkar

*University of North Carolina*

**1. Introduction.** S. N. Roy and Bargmann [3] used S. N. Roy's union-intersection method as the basis for providing tests and confidence intervals in the following cases:

i) $\mathbf{y}' = (y_1, \cdots, y_p) \sim N(\boldsymbol{\mu}', \Sigma)$, $H_o:\sigma_{ij} = 0$, $i \neq j$.
ii) $\mathbf{y}' \sim N(\boldsymbol{\mu}', \Sigma)$, but $\mathbf{y}'$ is partitioned into $k$ sets or blocks or sizes $p_1, \cdots, p_k$. $H_o:\Sigma_{ij} = 0$, $i \neq j$, where $\Sigma_{ij}$ is the covariance matrix between blocks $i$ and $j$.

J. Roy [1] considered the following additional cases:

iii) $Y:n \times p$, $(y_{1j}, \cdots, y_{pj}) \sim N(-, \Sigma)$, $j = 1, \cdots, n$, $EY = A\theta$. $A:n \times m$ has rank $r \leqq n - p$ and is known, $\theta$ is unknown. Let $\Phi = B\theta$ be estimable, $B:t \times m$. $H_o:\Phi = 0$.
iv) $(y_1, \cdots, y_p) \sim N(\boldsymbol{\mu}', \Sigma)$. $H_o:\Sigma = \Sigma_0$ (specified).
v) $(y_1, \cdots, y_p) \sim N(\boldsymbol{\mu}', \Sigma_1)$, $(x_1, \cdots, x_p) \sim N(\boldsymbol{\nu}', \Sigma_2)$, $H_o:\Sigma_1 = \Sigma_2$.

In this note we shall consider the following modification of (iii):

vi) $Y:n \times p$, $(y_{1j}, \cdots, y_{pj}) \sim N(-, \Sigma)$, $j = 1, \cdots, n$, $EY = A\theta$ (as in (iii)). $H_o:\sigma_{ij} = 0$, $i \neq j$.

**2. Step-down procedure to test $H_o$ in (vi).** In the notation of [1], denote the ith columns of the matrices $Y$ and $\theta$ by $\mathbf{y}_i$ and $\boldsymbol{\theta}_i$ respectively and write

$$Y_i = [\mathbf{y}_1, \cdots, \mathbf{y}_i], \qquad \theta_i = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_i]$$

and $\Sigma_i = (\sigma_{jk})$, $j, k = 1, \cdots, i$.

If $Y_i$ is fixed, the $n$ elements of $\mathbf{y}_{i+1}$ are distributed independently and normally with the same variance $\sigma_{i+1}^2$ and expectations given by

$$E(\mathbf{y}_{i+1} \mid Y_i) = A\boldsymbol{\eta}_{i+1} + Y_i\boldsymbol{\beta}_i, \tag{1}$$

where $\boldsymbol{\beta}_i':1 \times i$ is the row vector,

$$\boldsymbol{\beta}_i' = (\sigma_{1,i+1}; \cdots; \sigma_{i,i+1})\Sigma_i^{-1}, \tag{2}$$

and $\boldsymbol{\eta}_{i+1}:m \times 1$ is the column vector given by

$$\boldsymbol{\eta}_{i+1} = \boldsymbol{\theta}_{i+1} - \theta_i\boldsymbol{\beta}_i, \qquad i = 1, \cdots, p - 1. \tag{3}$$

We note that $H_o$ is true if and only if the hypothesis $H_i:\boldsymbol{\beta}_i = 0$ holds for all $i = 1, \cdots, p - 1$. Now the elements of the vectors $\boldsymbol{\beta}_i$ and $\boldsymbol{\eta}_{i+1}$ may be regarded

Received December 24, 1958; revised May 11, 1959.

as unknown parameters, and hence, when $Y_i$ is fixed, the hypothesis $H_i{:}\boldsymbol{\beta}_i = 0$ is a linear hypothesis in univariate analysis with the linear model given by (1).

We observe that rank $Y_i = i$, a.e. and rank $(A \vdots Y_i) = r + i$, a.e. Hence $\boldsymbol{\beta}_i$ is estimable and the hypothesis $H_i$ is testable. Let $\hat{\boldsymbol{\beta}}_i$ be the Gauss-Markov estimator of $\boldsymbol{\beta}_i$ in the conditional set-up. Denote the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_i$ by $\sigma_{i+1}^2 C_i$ where $C_i{:}i \times i$ is a positive-definite matrix. Let $s_i^2/n - r - i$ denote the usual error mean square giving an unbiased estimator of $\sigma_{i+1}^2$. Then, as in [1],

$$F_i \equiv \frac{(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)'C_i^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)/i}{s_i^2/(n - r - i)}, \qquad i = 1, \cdots, p - 1, \tag{4}$$

has the $F$ distribution with $i$ and $n - r - i$ degrees of freedom.

Thus the conditional distribution of $F_i$, given $Y_i$, does not involve $Y_i$ and hence does not involve $F_1, \cdots, F_{i-1}$. Therefore, the statistics $F_1, \cdots, F_{p-1}$ have independent $F$ distributions with degrees of freedom $i$ and $n - r - i$, $i = 1, \cdots, p - 1$ respectively.

For a preassigned constant $\alpha_i$, $0 < \alpha_i < 1$, let $f_i$ denote the upper 100 $\alpha_i$ percent point of the $F$ distribution with $i$ and $n - r - i$ degrees of freedom. Then the probability that simultaneously

$$F_i \leqq f_i, \qquad i = 1, \cdots, p - 1, \tag{5}$$

is equal to $\prod_{i=1}^{p-1} (1 - \alpha_i)$.

Since $H_o \leftrightarrow H_i{:}\boldsymbol{\beta}_i = 0$, $i = 1, \cdots, p - 1$, we utilize (4) and propose the following test procedure for $H_o$:
Accept $H_o$, if

$$u_i = \frac{\hat{\boldsymbol{\beta}}_i C_i^{-1}\hat{\boldsymbol{\beta}}/i}{s_i^2/n - r - i} \leqq f_i \qquad \text{for all } i = 1, \cdots, p - 1; \tag{6}$$

otherwise reject $H_o$.

To carry out the test one should first compute $u_1$. If $u_1 > f_1$, $H_o$ is rejected. If $u_1 \leqq f_1$, $u_2$ is computed. If $u_2 > f_2$, $H_o$ is rejected. If $u_2 \leqq f_2$, $u_3$ is computed and so on. The level of significance for this test is obviously $1 - \prod_{i=1}^{p-1} (1 - \alpha_i)$. One possibility is to choose $\alpha_1 = \cdots = \alpha_{p-1}$. We prefer choosing $\alpha$'s so that $f_1 = \cdots = f_{p-1}$, for reasons discussed in [3].

**3. Confidence bounds associated with the test.**

Now from (4), $F_i \leqq f_i \Rightarrow (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \leqq \lambda_{\max}(C_i)l_i^2 s_i^2$ where $l_i^2 = if_i/(n - r - i)$ and $\lambda_{\max}(C_i)$ is the maximum characteristic root of $C_i$. Hence, in view of (5), with a probability greater than $\prod_{i=1}^{p-1} (1 - \alpha_i)$,

$$(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \leqq \lambda_{\max}(C_i)l_i^2 s_i^2, \qquad i = 1, \cdots, p - 1. \tag{7}$$

Now (7) implies

$$\mathbf{a}_i'\hat{\boldsymbol{\beta}}_i - l_i s_i \lambda_{\max}^{1/2}(C_i) \leqq \mathbf{a}_i'\boldsymbol{\beta}_i \leqq \mathbf{a}_i'\hat{\boldsymbol{\beta}}_i + l_i s_i \lambda_{\max}^{1/2}(C_i) \tag{8}$$

for all non-null $\mathbf{a}_i:i \times 1$ such that $\mathbf{a}_i'\mathbf{a}_i = 1$, $(i = 1, \cdots, p-1)$. This again implies

$$(9)\qquad \begin{aligned}(\hat{\boldsymbol\beta}_i'\hat{\boldsymbol\beta}_i)^{1/2} - l_i s_i \lambda_{\max}^{1/2}(C_i) &\leqq (\boldsymbol\beta_i'\boldsymbol\beta_i)^{1/2} \\ &\leqq (\hat{\boldsymbol\beta}_i'\hat{\boldsymbol\beta}_i)^{1/2} + l_i s_i \lambda_{\max}^{1/2}(C_i), \quad i = 1, \cdots, p-1.\end{aligned}$$

Thus (9) holds with probability greater than $\prod_{i=1}^{p-1}(1-\alpha_i)$. We may obtain partial statements by choosing some elements of $\mathbf{a}_i$ in (8) to be zero. Thus we have the simultaneous confidence bounds given by (9) for all possible subsets of $\boldsymbol\beta_i$ for all $i = 1, \cdots, p-1$ with the confidence coefficient greater than $\prod_{i=1}^{p-1}(1-\alpha_i)$.

**4. Remarks.**

(a) It will be easily seen that when $Y$ represents a random sample of size $n$ from $N(\boldsymbol\mu, \Sigma)$, (1) takes the form

$$E(y_{i+1,k} \mid Y_i) = \mu_{i+1} + \sum_{j=1}^{i} \beta_{ij}(y_{jk} - \mu_j),$$

where $\mathbf{y}_i' = (y_{i1}, \cdots, y_{in})$ and $\boldsymbol\beta_i' = (\beta_{i1}, \cdots, \beta_{ii})$, $i = 1, \cdots, p-1$.

If we write $s_{ij} = \sum_{k=1}^{n}(y_{ik} - \bar y_i)(y_{jk} - \bar y_j)$ and $S_i = (s_{jk})$, $j, k = 1, \cdots, i$, then it is well-known that

$$\hat{\boldsymbol\beta}_i = S_i^{-1}\begin{pmatrix} s_{i+1,1} \\ \vdots \\ s_{i+1,i}\end{pmatrix} = \mathbf{b}_i, \qquad C_i = S_i^{-1}$$

and

$$s_i^2 = s_{i+1,i+1} - (s_{i+1,1}; \cdots; s_{i+1,i})S_i^{-1}(s_{i+1,1}; \cdots; s_{i+1,i})',$$

so that

$$u_i = \frac{\mathbf{b}_i'S_i\,\mathbf{b}_i/i}{s_i^2/n-1-i} = \frac{r^2_{i+1.1,\cdots,i}}{1-r^2_{i+1.1,\cdots,i}}\,\frac{n-1-i}{i},$$

where $r_{i+1.1,\cdots,i}$ denotes the multiple correlation coefficient of $(i+1)$ with $(1, \cdots, i)$, thus giving as a special case the test procedure already obtained in [3]. This is, of course, as it should be.

(b) In this model, as in (iii), it is of interest to investigate whether the test of the usual multivariate linear hypothesis of the type

$$(10)\qquad H_0':\Phi = B\theta = 0,$$

where $\Phi$ is estimable, and the above test of independence are quasi-independent (see e.g. Roy [2]). As shown in [1], the step-down test procedure for (10) gives, when $Y_i$ is fixed,

$$(11)\qquad F_i' \equiv \frac{(\hat{\boldsymbol\phi}_{i+1} - \boldsymbol\phi_{i+1})'D_{i+1}^{-1}(\hat{\boldsymbol\phi}_{i+1} - \boldsymbol\phi_{i+1})/t}{s_i^2/n-r-1}, \quad i = 0, 1, \cdots, p-1$$

where $\Phi_{i+1} = B\mathbf{n}_{i+1}$ and the variance-covariance matrix of $\hat{\phi}_{i+1}$ is $D_{i+1}\sigma^2_{i+1}$. $F_i$ given by (4) and $F'_i$ given by (11) are, for fixed $Y_i$, quasi-independent if the numerators, which are marginally distributed as $\chi^2_i\sigma^2_{i+1}/i$ and $\chi^2_t\sigma^2_{i+1}/t$ respectively, are independent. It can be easily verified that $\chi^2_i$ and $\chi^2_t$ are not independent and hence the tests for $H_o$ and $H'_o$ are not quasi-independent. It may be noted that, when $Y_i$ is fixed, the test of $\beta_i = 0$ is like testing significance of regression, as seen from (1), while the test of $\Phi_{i+1} = 0$ is like covariance-analysis.

(c) We may consider extension of (vi) to blocks, as in (ii), and test

$$H_o: \Sigma_{ij} = 0,$$

as pointed out by the referee. It is easy to check that a similar step-down procedure with respect to blocks will result in $k - 1$ independent tests in multivariate analysis of variance of the same general structure as in [1] and [3].

**5. Acknowledgement.** I am indebted to Professor S. N. Roy for suggesting this problem and to the referee for suggesting improvements in structure and exposition.

## REFERENCES

[1] J. Roy, "Step-down procedure in multivariate analysis," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1177–1187.

[2] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.

[3] S. N. Roy and R. E. Bargmann, "Tests of multiple independence and the associated confidence bounds," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 491–503.

## NON-MARKOVIAN PROCESSES WITH THE SEMIGROUP PROPERTY

By William Feller[1]

*Princeton University*

**1. Introduction.** Every $N \times N$ stochastic matrix $P$ defines the transition probabilities of a Markovian process with positive discrete time parameter. Its $n$-step transition probabilities satisfy the Chapman-Kolmogorov, or semigroup, relation $P^{n+m} = P^n P^m$. We shall show that *for $N \geqq 3$ there exist non-Markovian processes with $N$ states whose transition probabilities satisfy the same equation.*[2] All elements of $P$ will equal $N^{-1}$. The process may be chosen *strictly stationary.* A simple modification leads to non-Markovian processes with *continuous time* parameter and the semigroup property with $N$ states or a continuum of states.

The triviality of the following example should not obscure the interest of the problem concerning the existence of non-Markovian processes satisfying the Chapman-Kolmogorov equation. As so many other basic problems in probability theory, it has been formulated by P. Lévy who with his usual ingenuity gave the first counter-example to the obvious conjecture.

**2.** Let $\mathfrak{P}$ be the sample space whose points $(x^{(1)}, \cdots, x^{(N)})$ are the random permutations of $(1, 2, \cdots, N)$ each carrying probability $1/N!$ Let $\mathfrak{R}$ be the set of the $N$ points $(x^{(1)}, \cdots, x^{(N)})$ such that $x^{(i)} = \nu$ for all $1 \leqq i \leqq N$ where $\nu$ is a fixed integer $1 \leqq \nu \leqq N$; each point of $\mathfrak{R}$ carries probability $1/N$. Finally, Let $\mathfrak{S}$ be the mixture of $\mathfrak{P}$ and $\mathfrak{R}$ with $\mathfrak{P}$ carrying weight $1 - N^{-1}$ and $\mathfrak{R}$ weight $N^{-1}$.

More formally, $\mathfrak{S}$ contains the $N! + N$ arrangements $(x^{(1)}, x^{(2)}, \cdots, x^{(N)})$ which represent either a premutation of $(1, 2, \cdots, N)$ or the $N$-fold repetition of an integer $\nu$, $1 \leqq \nu \leqq N$. To each point of the first class we attribute probability $(1 - N^{-1})(N!)^{-1}$, to each point of the second class probability $N^{-2}$.

Then clearly

$$(1) \qquad P\{x^{(i)} = \nu\} = N^{-1}, \qquad P\{x^{(i)} = \nu, \quad x^{(j)} = \mu\} = N^{-2}$$

for all $i \neq j$. Thus all transition probabilities are equal:

$$(2) \qquad P\{x^{(i)} = \nu \mid x^{(j)} = \mu\} = N^{-1}.$$

Given, say, that $x^{(1)} = 1$, $x^{(2)} = 1$ the probability that $x^{(3)} \neq 1$ is zero, and hence the process is non-Markovian.

**3.** To extend the process to all integral values of the time parameter consider, in the usual manner, a double infinity of independent repetitions of the described

---

Received February 4, 1959.

[1] Research supported by the Office of Ordnance Research U. S. Army Contract DA-36-034-ORD-2001.

[2] [Added in proof.] D. Blackwell has pointed out to me that the variables of our process represent a sequence of *random variables which are pairwise independent without being mutually independent.*

sample space. In other words, we consider the product space $\cdots \mathfrak{S} \times \mathfrak{S} \times \mathfrak{S} \cdots$ with product measure; its points are the doubly infinite sequences $x = \{x^{(i)}\}$ such that for each integer $r$ the $N$-dimensional block $(x^{(rN+1)}, x^{(rN+2)}, \cdots, x^{(r+1)N})$ represents the projection of $x$ onto a coordinate space $\mathfrak{S}$. This represents a non-Markovian process with the stationary transition probabilities (2). However, the process itself is not stationary in view of the periodicity modulo $N$.

To obtain a *stationary* process of the same type it suffices to introduce $N$ replicas of our process with time shifts $0, 1, 2, \cdots, N - 1$ and define a new process as their mixture with equal weights.

**4.** To construct a process of a *similar character defined for all* $t \geq 0$ consider the above discrete process and a Poisson process $\{N(t)\}$ with mean $t$ independent of it. Define a new process by

$$x(t) = x^{(N(t))}. \tag{3}$$

Its absolute probabilities are given by

$$P\{x(t) = \nu\} = \sum_{i=0}^{\infty} P\{N(t) = i\} \cdot P\{x^{(i)} = \nu\} = N^{-1}. \tag{4}$$

The joint probabilities for $0 \leq s < t$ are calculated in like manner, but the possibility that $N(t) = N(s)$ makes it necessary to consider separately the cases $\nu = \mu$ and $\nu \neq \mu$. Clearly

$$\begin{aligned} P_{\nu\mu}(t) &= P\{x(s + t) = \mu \mid x(s) = \nu\} \\ &= N^{-1}(1 - e^{-t}), \qquad \text{if } \nu \neq \mu, \\ P_{\nu\nu}(t) &= e^{-t} + N^{-1}(1 - e^{-t}) \end{aligned} \tag{5}$$

and a simple calculation shows that

$$P_{\nu\mu}(s + t) = \sum_{\lambda=1}^{N} P_{\nu\lambda}(s) P_{\lambda\mu}(t) \tag{6}$$

even though the process is clearly non-Markovian.

Finally, one might replace the $N$ states by $N$ intervals and an appropriate motion within them.

## REFERENCE

[1] P. Lévy, "Exemples de processus pseudo-Markoviens," *Comptes Rendus Académie Sciences*, Paris, Vol. 228, (1949), pp. 2004–2006.

# SUCCESSIVE RECURRENCE TIMES IN A STATIONARY PROCESS

By Shu-Teh Chen Moy

*Wayne State University*

Let $X_0, X_1, X_2, \cdots$ be a stationary sequence of random variables. Let $B$ be any linear Borel set for which $P(X_0 \varepsilon B) > 0$. We are concerned with the successive recurrence times $\nu_1, \nu_2, \cdots$ of $B$; their time averages and their expectations. Without loss of generality, we shall assume the basic probability space $\Omega$ to be the collection of all sequences $\omega = \{\cdots, x_{-1}, x_0, x_1, \cdots\}$ and $X_n$ to be the coordinate variables, i.e., $X_n(\omega) = x_n$. Let $T$ be the shift transformation. The $n$th coordinate of $T\omega$ is the $(n + 1)$th coordinate of $\omega$. Then $T$ is 1–1 and preserves the probability measure $P$. For any

$$\omega = \{\cdots x_{-1}, x_0, x_1, \cdots\},$$

if there are infinitely many positive integers $n$ with $x_n \varepsilon B$, let

$$\nu_1, \nu_1 + \nu_2, \cdots, \nu_1 + \nu_2 + \cdots + \nu_k, \cdots$$

be the successive positive integers for which $x_{\nu_2+\cdots-\nu_k} \varepsilon B$. If there are finitely many, say $K$, positive integers $n$ with $x_n \varepsilon B$, define $\nu_1, \cdots, \nu_K$ as before but define $\nu_{K+1} = \nu_{K+2} = \cdots = \infty$. In this paper, Theorem 1 is concerned with the time average of the successive recurrence times, the $\nu$'s. In Theorem 2 the successive recurrence times are proved to be stationary given $X_0 \varepsilon B$. Theorem 3 may be considered as a generalization of a theorem of M. Kac in which he proved the formula (7) for the first recurrence time $\nu_1$ ([2], pp. 1006).

Theorem 1: *For almost all* $\omega$

$$\lim_{k\to\infty} \frac{\nu_1(\omega) + \cdots + \nu_k(\omega)}{k} \tag{1}$$

*exists. The limit may be finite or infinite. It is finite for almost all* $\omega \varepsilon E$ *where* $E = [X_0 \varepsilon B]$. *In particular, if* $T$ *is ergodic, the limit is equal to* $1/P(E)$ *with probability one.*

Proof: Let $I_E$ be the indicator function of $E$, i.e., $I_E(\omega) = 1$ if $\omega \varepsilon E$ and $I_E(\omega) = 0$ if $\omega \not\varepsilon E$. By the ergodic theorem, for almost all $\omega$

$$\lim_{k\to\infty} \frac{I_E(T\omega) + \cdots + I_E(T^n\omega)}{n} = f(\omega), \tag{2}$$

where $f(\omega) > 0$ for almost all $\omega \varepsilon E$. If $T$ is ergodic $f(\omega) \equiv P(E)$.

In fact, $[I_E(T\omega) + \cdots + I_E(T^n\omega)]n^{-1}$ is the relative frequency of occurrence of $B$. If the limit of the relative frequency, as $n \to \infty$, is positive, $B$ must occur infinitely often; therefore, $\nu_1(\omega), \nu_2(\omega), \cdots$ are all finite. Thus all successive recurrence times are finite for almost all $\omega \varepsilon E$. In particular, if $T$ is ergodic, they are all finite for almost all $\omega \varepsilon \Omega$.

Received October 14, 1958; revised June 28, 1959.

Let $\Omega'$ be the collection of all $\omega$ for which $\nu_1(\omega), \nu_2(\omega), \cdots$ are all finite. Let $\omega \varepsilon \Omega'$. For any positive integer $k$, let $n_k = \nu_1(\omega) + \cdots + \nu_k(\omega)$. Then

$$\frac{\nu_1(\omega) + \cdots + \nu_k(\omega)}{k} = \frac{n_k}{I_E(T\omega) + \cdots + I_E(T^{n_k}\omega)}.$$

Therefore, for almost all $\omega \varepsilon \Omega'$, there exists the limit

$$\lim_{k\to\infty} \frac{\nu_1(\omega) + \cdots + \nu_k(\omega)}{k} = \frac{1}{f(\omega)}. \tag{3}$$

The limit is finite or infinite according as $f(\omega) > 0$ or $f(\omega) = 0$. If $\omega \not\varepsilon \Omega'$ then there is a positive integer $K$ for which $\nu_{K+1}(\omega) = \nu_{K+2}(\omega) = \cdots = \infty$. Therefore

$$\lim_{k\to\infty} \frac{\nu_1(\omega) + \cdots + \nu_k(\omega)}{k} = \infty.$$

It is clear that $I_E(T\omega) + \cdots + I_E(T^n\omega) \leqq K$ for all $n$ and that

$$\lim_{n\to\infty} \frac{I_E(T\omega) + \cdots + I_E(T^n\omega)}{n} = 0.$$

Therefore (3) again holds true. Hence (3) holds true with probability one. If $T$ is ergodic, $1/f(\omega) \equiv 1/p(E)$.

Let $P_E$, $E = [X_0 \varepsilon B]$, be the conditional probability measure given $X_0 \varepsilon B$, i.e., for any measurable set $F$,

$$P_E(F) = P(E \cap F)/P(E). \tag{4}$$

Then $\nu_1, \nu_2, \cdots$ are finite valued with probability one under the probability measure $P_E$.

THEOREM 2. *The random variables $\nu_1, \nu_2, \cdots$ are stationary under the conditional probability measure $P_E$, i.e.,*

$$P_E(\nu_1 = i_1, \cdots \nu_k = i_k) = P_E(\nu_{m+1} = i_1, \cdots \nu_{m+k} = i_k) \tag{5}$$

*for any positive integers, $m$, $k$, and any $k$-tuple of positive integers, $(i_1, \cdots i_k)$.*

PROOF: We shall proceed by induction on the integer $m$. Let $F_{i_1}, \cdots i_k = [\nu_1 = i_1, \cdots \nu_k = i_k]$, and let $E' = \Omega - E$. Then

$$\begin{aligned}
P_E[\nu_2 = i_1, \cdots \nu_{k+1} = i_k] &= \sum_{n=1}^{\infty} P_E[\nu_1 = n, \nu_2 = i_1, \cdots \nu_{k+1} = i_k] \\
&= \sum_{n=1}^{\infty} P_E[T^{-1}E' \cap \cdots \cap T^{-(n-1)}E' \cap T^{-n}E \cap T^{-n}F_{i_1,\cdots i_k}] \\
&= \frac{1}{P(E)} \sum_{n=1}^{\infty} P[E \cap T^{-1}E' \cap \cdots \cap T^{-(n-1)}E' \cap T^{-n}E \cap T^{-n}F_{i_1,\cdots,i_k}] \\
&= \frac{1}{P(E)} \sum_{n=1}^{\infty} P[T^nE \cap T^{(n-1)}E' \cap \cdots \cap TE' \cap E \cap F_{i_1,\cdots i_k}] \\
&= \frac{1}{P(E)} P\left[\left(\bigcup_{n=1}^{\infty} T^nE\right) \cap E \cup F_{i_1,\cdots,i_k}\right]
\end{aligned}$$

The Poincaré recurrence theorem ([1], pp. 10) asserts that

$$P\left[\left(\bigcup_{n=1}^{\infty} T^n E\right) \cap E\right] = P(E).$$

Hence

$$P_E[\nu_2 = i_1, \cdots, \nu_{k+1} = i_k] = \frac{1}{P(E)} P[E \cap F_{i_1,\cdots,i_k}] = P_E[F_{i_1}, \cdots i_k]$$
$$= P_E[\nu_1 = i_1, \cdots \nu_k = i_k].$$

Hence (5) is true for $m = 1$ and any $k$ and any $k$-tuple $(i_1, \cdots i_k)$. Now assume that (5) holds true for all $m \leqq M$.

$$PE[\nu_{M+2} = i_1, \cdots \nu_{M+1+k} = i_k] = \sum_{n=1}^{\infty} P_E[\nu_{M+1} = n, \nu_{M+2} = i_1, \cdots, \nu_{M+1+k} = i_k]$$
$$= \sum_{n=1}^{\infty} P_E[\nu_1 = n, \nu_2 = i_1, \cdots, \nu_{1+k} = i_k] = P_E[\nu_2 = i_1, \cdots, \nu_{k+1} = i_k]$$
$$= P_E[\nu_1 = i_1, \cdots, \nu_k = i_k].$$

Hence (5) is true for all $m$.

THEOREM 3: *Let $f(\omega)$ be defined by* (2), *i.e., $f(\omega)$ is the limit, as $n \to \infty$, of the relative frequency of occurrence of $B$. Then, for any $k$,*

$$\int \nu_k(\omega) P_E(d\omega) = \int \frac{1}{f(\omega)} P_E(d\omega). \tag{6}$$

*The conditional expectation of the kth recurrence time given $X_0 \varepsilon B$ is finite if and only if $1/f(\omega)$ is integrable with respect to $P_E$. In particular, if the shift transformation $T$ is ergodic, then*

$$\int \nu_k(\omega) P_E(d\omega) = \frac{1}{P(E)}. \tag{7}$$

PROOF. By Theorem 1, the set of all $\omega$ such that

$$\lim_{k\to\infty} \frac{\nu_1(\omega) + \cdots + \nu_k(\omega)}{k} = \frac{1}{f(\omega)}$$

has $P_E$ measure 1. Since the process $\nu_1, \nu_2, \cdots$ is stationary under $P_E$ by Theorem 2, the conditional expectations $\int \nu_k(\omega) P_E(d\omega)$ are the same for all $k$. If $\int \nu_k(\omega) P_E(d\omega) < \infty$, since $\{\nu_k\}$ is stationary, (6) follows easily from the ergodic theorem. If $\int \nu_k(\omega) P_E(d\omega) = \infty$, let

$$\nu_k^N(\omega) = \begin{cases} \nu_k(\omega), & \nu_k(\omega) \leqq N, \\ N, & \text{otherwise.} \end{cases}$$

Then the process $\nu_1^N, \nu_2^N, \cdots$ is again stationary under $P_E$, and therefore the set of $\omega$ for which $\lim_{K\to\infty} (\nu_1^N(\omega) + \cdots + \nu_k^N(\omega)/k$ exists has $P_E$ measure 1. Let $g_N(\omega)$ be the limit. We have

$$\int \nu_k^N(\omega) P_E(d\omega) = \int g_N(\omega) P_E(d\omega).$$

But $g_N(\omega) \leqq 1/f(\omega)$, hence

$$\int \nu_k^N(\omega) P_E(d\omega) \leqq \int \frac{1}{f(\omega)} P_E(d\omega).$$

Since

$$\lim_{N\to\infty} \int \nu_k^N(\omega) P_E(d\omega) = \int \nu_k(\omega) P(d\omega) = \infty,$$

hence

$$\int \frac{1}{f(\omega)} P_E(d\omega) = \infty,$$

and again (7) is true.

## REFERENCES

(1) P. R. Halmos, *Lectures on Ergodic Theory*, The Mathematical Society of Japan, Tokyo, 1956.

(2) M. Kac, "On the Notion of Recurrence in Discrete Processes," *Bull. Amer. Math. Soc.*, Vol. 53 (1947), pp. 1002–1010.

# ON THE MUTUAL INDEPENDENCE OF CERTAIN STATISTICS

BY C. G. KHATRI

*University of Baroda, India*

**1. Summary and introduction.** The results of this note yield the mutual independence of certain matrices, characteristic roots, Hotelling's $T^2$ or Mahalanobis' $D^2$ statistics, and C. R. Rao's $R$ statistic. The result concerning the mutual independence of certain of Hotelling's $T^2$ statistics has been proved by K. S. Rao [1]. The results mentioned here occur (by implication and as a by-product) in [5], [6] and [7] in the course of investigations on some specific problems in statistical inference but are not explicitly stated. These results can be utilised in statistical inference and especially in simultaneous tests and simultaneous confidence interval estimation, and also in other problems.

**2. Certain known results.**

(2.1) Let $X:p \times n(n \geqq p)$ be a matrix of $p$ rows and $n$ columns. (A column vector $x:p \times 1$ is denoted as $\underset{\sim}{x}:p \times 1$). Let $X$ have a distribution $f(X)$. Then $XX'$ is symmetric positive definite.

(2.2) If $S:p \times p$ is symmetric positive definite, then $S = \tilde{T}\tilde{T}'$, where $\tilde{T}$ is a triangular matrix with zero's above the principal diagonal, $t_{ii} > 0$, and $t_{ji} = | A_{ji} | / \sqrt{|A_{ii}| \cdot | A_{i-1,i-1} |}$, where

$$A_{ji} = \begin{pmatrix} s_{11} & \cdots & s_{1i} \\ \cdots & \cdots & \cdots \\ s_{i-1,1} & \cdots & s_{i-1,i} \\ s_{j1} & \cdots & s_{ji} \end{pmatrix}, \qquad j \geqq i.$$

(See [2].)

(2.3) The roots of $XX'$ are the roots of $X'X$ except for some zero roots. (See [5], [7].)

(2.4) If $X_j$ and $Y_j$ are transformed to $X_{j+1}$ and $Y_{j+1}$ respectively, $j = 1, \cdots k$, by the following matrix transformations,

$$X_j = F_j(X_{j+1}, Y_j) \text{ and } Y_j = G_j(X_{j+1}, Y_{j+1}) \quad (j = 1, 2, \cdots k),$$

then the Jacobian of the transformation $X_1$ to $X_{k+1}$ and $Y_1$ to $Y_{k+1}$ is

$$J(X_1, Y_1; X_{k+1}, Y_{k+1}) = \textstyle\prod_{j=1}^{k} J(X_j; X_{j+1}) J(Y_j; Y_{j+1}).$$

(See [2], [5], [7].)

(2.5) The Jacobian of the transformation $X = AYB$ ($X:p \times q$, $Y:p \times q$, $A:p \times p$, $B:q \times q$) is $J(X; Y) = | A |^q | B |^p$. (See [3].)

(2.6) The Jacobian of the transformation $S = GRG'$ ($S$, $R:p \times p$ symmetric matrices, $G:p \times p$) is $J(S; R) = | G |^{p+1}$. (See [3].)

Received December 16, 1957; revised March 9, 1959.

(2.7) The Jacobian of the transformation $S = R^{-1}(S, R{:}p \times p$ symmetric matrices) is $J(S; R) = |R|^{-(p+1)}$. (See [3].)

$$\int_{\substack{S \leq XX' \leq S+(ds_{ij}) \\ X:p\times n\,(n\geq p)}} \cdots \int dX = \pi^{[2pn-p(p-1)]/4} \prod_{i=1}^{p} \left\{ \Gamma\left(\frac{n-i+1}{2}\right) \right\}^{-1} \cdot |S|^{(n-p-1)/2}\, dS, \tag{2.8}$$

where $dQ$ is the product of the differentials of the variables. (See [4], [5], [7].)

**3. Theorem I.** *If $S{:}p \times p$ and $X{:}p \times q$ are independently distributed as Wishart $(n, p; \Sigma; S)$ and $MN(0, \Sigma)$ respectively (where $MN(0, \Sigma)$ is called multivariate normal and has a density which is a multiple of $\exp[-(\operatorname{tr} \Sigma^{-1}XX')/2]$), then $S_1 = S + XX'$ and $Z = \tilde{T}^{-1}X$ (where $S = \tilde{T}\tilde{T}'$) are independent with distributions given by Wishart $(n + q, p; \Sigma; S_1)$ and the density $c(I - ZZ')^{(n-p-1)/2}$ respectively. Here*

$$c = \prod_{i=1}^{p} \Gamma\left(\frac{n+q-i+1}{2}\right) \left\{ \Gamma\left(\frac{n-i+1}{2}\right) \right\}^{-1} \pi^{-pq/2}.$$

Proof. Transform by the relations $S = S_1 - XX'$ and $X = \tilde{T}Z$, where $S_1 = \tilde{T}\tilde{T}'$; then by (2.4) and (2.5), the Jacobian of the transformation is $|\tilde{T}|^q = |S_1|^{q/2}$; also $|S| = |S_1|\,|I - ZZ'|$.

Since $S$ and $X$ are independently distributed, we can easily see that the joint distribution of $S_1$ and $Z$ is

$$f(S_1, Z) = f_1(S_1) f_2(Z), \tag{1}$$

where

$$\begin{aligned} f_1(S_1) &= \text{Wishart } (n + q, p; \Sigma; S_1), \\ f_2(Z) &= c\,|I - ZZ'|^{(n-p-1)/2}, \end{aligned} \tag{2}$$

$c$ being the same as defined in the Theorem.

Corollary 1. *If $S{:}p \times p$, $X_i{:}p \times q_i$, $i = 1, 2, \cdots m$, are independently distributed as Wishart $(n, p; \Sigma; S)$ and $MN_i(0, \Sigma)$ respectively, then*

*$S_m = S + \sum_{j=1}^{m} X_jX_j'$ and $Z_i = \tilde{T}_i^{-1}X_i$ (where $S_i = \tilde{T}_i\tilde{T}_i' = S + \sum_{j=1}^{i} X_jX_j'$)*

*are mutually independent and distributed with the respective densities Wishart $(n + e_m, p; \Sigma; S_m)$ and*

$$\pi^{-pq_i/2} \prod_{j=1}^{p} \cdot\Gamma\left(\frac{n+e_i-j+1}{2}\right) \left\{ \Gamma\left(\frac{n+e_{i-1}-j+1}{2}\right) \right\}^{-1} |I - Z_iZ_i'|^{(n+e_{i-1}-p-1)/2},$$

$$e_i = \textstyle\sum_{j=1}^{i} q_j, \qquad i = 1, 2, \cdots m. \tag{3}$$

The proof can be obtained in a similar manner as above.

Corollary 2. *In Corollary 1, suppose $p \geq q_i$. Then the distribution of*

$V_i = Z_i' Z_i = X_i'(S + \sum_{j=1}^{i} X_j X_j')^{-1} X_i$ *is given by*

$$(4) \qquad c_1 \mid V_i \mid^{(p-q_i-1)/2} \mid I - V_i \mid^{(n+e_{i-1}-p-1)/2},$$

$$c_1 = \prod_{j=1}^{q_i} \cdot \Gamma\left(\frac{n + e_i - j + 1}{2}\right) \left\{ \Gamma\left(\frac{p - j + 1}{2}\right) \Gamma\left(\frac{n + e_i - j - p + 1}{2}\right) \right\}^{-1} \pi^{-q_i(q_i-1)/4}.$$

Proof. The distribution of $V_i$ can at once be obtained by the use of the integral (2.8) in formula (3).

Note that the distribution of $W_i = V_i(I - V_i)^{-1} = X_i'(S + \sum_{j=1}^{i-1} X_j X_j')^{-1} X_i$ is obtained by using the Jacobian (2.6) for $(I - V_i) = (I + W_i)^{-1}$ and we find it to be

$$(5) \qquad c_1 \mid W_i \mid^{(p-q_i-1)/2} \mid I + W_i \mid^{-(n+e_i)/2}.$$

Similarly if $p \leqq q_i$, the distribution of $Z_i Z_i'$ and $Z_i Z_i'(I - Z_i Z_i')^{-1}$ can be obtained from (3).

Corollary 3. *If in Corollary* 1, $q_i \equiv 1$ *for all i, then*

$$T_i^2 = x_i' (S + \sum_{j=1}^{i-1} x_j x_j')^{-1} x_i, \qquad i = 1, 2 \cdots m$$

*and*

$$S_m = S + \sum_{j=1}^{m} x_j x_j',$$

*are mutually independent with distributions given by*

$$\text{const.} (T_i^2)^{(p-2)/2} (1 + T_i^2)^{-(n+i)/2}, \qquad i = 1, 2, \cdots m$$

*and Wishart* $(m + n, p; \Sigma; S_m)$.

Corollary 4. *If* $S:p \times p$, $X_i : p \times q_i$, $i = 1, 2, \cdots m$ *are independently distributed as Wishart* $(n, p; \Sigma; S)$ *and* $MN_i(0, \Sigma)$, *then the characteristic roots of* $A_i(S + \sum_{j=1}^{i-1} A_j)^{-1}$ *are distributed independently of those of* $S_m = S + \sum_{j=1}^{m} A_j$ (*where* $A_j = X_j X_j'$).

Proof. By Corollary 1, we have $Z_i$ and $S_m$, $i = 1, 2, \cdots m$, independently distributed and hence roots of $Z_i Z_i'$ and $S_m$ are independently distributed; i.e. by the application of (2.3) we have the corollary.

Theorem 2. *Let* $S:p \times p$, $x:p \times 1$ *be independently distributed as Wishart* $(n, p; \Sigma; S)$ *and multivariate normal* $(\mu, \Sigma)$, $\mu \neq 0$, *respectively, and let*

$$x: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{p-q}^{q}, \quad S: \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}_{p-q}^{q}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{p-q}^{q},$$

with column blocks of sizes $q$ and $p-q$,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}_{p-q}^{q}, \qquad \Delta_q^2 = \mu_1' \Sigma_{11}^{-1} \mu_1 \text{ and } \Delta_p^2 = \mu' \Sigma^{-1} \mu.$$

*If* $\Delta_p^2 = \Delta_q^2$, *then* $R = (1 + \underline{x}_1'S_{11}^{-1}\underline{x}_1)/(1 + \underline{x}'S^{-1}\underline{x})$ *and* $S_1 = S + \underline{x}\underline{x}'$ *are independently distributed and their densities are constant* $R^{(n-p-1)/2}(1 - R)^{(p-q-2)/2}$ *and non-central Wishart* $(n + 1, p, 1; \Sigma, \underline{\mu}; S_1)$ *with* 1 (1 *is trivially the rank of* $\underline{\mu}$) *non-central parameter* (for a discussion of the $R$ statistic, see [8]).

Proof. Let $2\Sigma = \tilde{B}\tilde{B}'$. Make the transformations $V_0 = \tilde{B}^{-1}S\tilde{B}'^{-1}$, $y = \tilde{B}^{-1}\underline{x}$, $V_1 = V_0 + yy'$ and $\tilde{T}^{-1}y \equiv \underline{w}$ where $V_1 = \tilde{T}\tilde{T}'$. By (2.5) and (2.6) the Jacobian of the transformation is $| \tilde{B} |^{p+2}| \tilde{T} | = | \Sigma |^{(p+2)/2}| V_1 |^{1/2}$. Also

$$| S | = | V_1 |.| I - \underline{w}\underline{w}' | = | \Sigma |.| V_1 |.(1 - \underline{w}'\underline{w}).$$

Let

$$\underline{\delta} = \tilde{B}^{-1}\underline{\mu} = \begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix}_{p-q}^{q}, \quad \tilde{B} = \begin{pmatrix}\tilde{B}_1 & \cdot\\ B_2 & \tilde{B}_3\end{pmatrix}_{p-q}^{q}, \quad \underline{w} = \begin{pmatrix}\underline{w}_1\\ \underline{w}_2\end{pmatrix}_{p-q}^{q}, \quad \tilde{T} = \begin{pmatrix}\tilde{T}_1 & \cdot\\ T_2 & \tilde{T}_3\end{pmatrix}_{p-q}^{q}.$$
$$\qquad\qquad\qquad\qquad q \quad p-q \qquad\qquad\qquad\qquad\qquad q \quad p-q$$

Then we get $2\Sigma_{11} = \tilde{B}_1\tilde{B}_1'$, $\underline{\delta}_1'\underline{\delta}_1 = \underline{\mu}_1'\Sigma_{11}^{-1}\underline{\mu}_1/2 = \Delta_q^2/2$ and $\underline{\delta}'\underline{\delta} = \Delta_p^2/2$. Hence if $\Delta_p^2 = \Delta_q^2$, we have

$$\underline{\delta}_2'\underline{\delta}_2 = 0 \text{ i.e. } \underline{\delta}_2 = \underline{0}. \tag{7}$$

Hence we can write down the joint distribution of $V_1$ and $\underline{w}$ as

$$f(V_1, \underline{w}) = \frac{(1 - \underline{w}'\underline{w})^{(n-p-1)/2} | V_1 |^{(n-p)/2} \exp(-\text{tr } V_1 + 2\underline{\delta}_1'\tilde{T}_1\underline{w}_1 - \Delta_q^2/2)}{\pi^{p(p+1)/4} \prod_{i=1}^{p} \Gamma\left(\frac{n-i+1}{2}\right)}. \tag{8}$$

Apply the transformation $\underline{z}_2 \equiv \underline{w}_2/\sqrt{(1 - \underline{w}_1'\underline{w}_1)}$. The Jacobian is $(1 - \underline{w}_1'\underline{w}_1)^{(p-q)/2}$. Hence equation (8) is

$$f(V_1, \underline{w}_1, \underline{z}_2) \equiv f_1(V_1, \underline{w}_1)f_2(\underline{z}_2) \tag{9}$$

where

$$f_1(V_1, \underline{w}_1) = c_2(1 - \underline{w}_1'\underline{w}_1)^{(n-q-1)/2}| V_1 |^{(n-p)/2} \exp(\text{-tr } V_1 + 2\underline{\delta}_1'\tilde{T}_1\underline{w}_1 - \Delta_q^2/2), \tag{10}$$

$$c_2 \text{ being } \pi^{-(2q+p^2-p)/4}\, \Gamma\left(\frac{n-p+1}{2}\right)\left\{\Gamma\left(\frac{n-q+1}{2}\right)\right\}^{-1} \prod_{i=1}^{p}\left\{\Gamma\left(\frac{n-i+1}{2}\right)\right\}^{-1}$$

and

$$f_2(\underline{z}_2) = \pi^{-(p-q)/2}\,\Gamma\left(\frac{n-q+1}{2}\right)\left\{\Gamma\left(\frac{n-p+1}{2}\right)\right\}^{-1}(1 - \underline{z}_2'\underline{z}_2)^{(n-p-1)/2}. \tag{11}$$

Hence $\underline{z}_2$ and $V_1$ are independently distributed; i.e. $V_1$ and $\underline{z}_2'\underline{z}_2$ are independently distributed. Note that $\underline{z}_2'\underline{z}_2 = 1 - R = (\underline{w}'\underline{w} - \underline{w}_1'\underline{w}_1)/(1 - \underline{w}_1'\underline{w}_1)$ for $1 - \underline{w}'\underline{w} = 1 - y'V_1^{-1}y = 1/(1 + y'V_0^{-1}y)$. Hence the distribution of $R$ can be easily obtained by the use of the integral (2.8) in (11). By integrating over $\underline{w}_1$ in (10), it can be easily shown that the distribution of $V_1$ (and so of $S_1$) is non-central Wishart with $(n + 1)$ d.f. and one (1 is the rank of $\underline{\mu}$) non-central parameter.

Corollary 5. *Let* $S{:}p \times p$, $\underline{x}_i{:}p \times 1$, $i = 1, 2, \cdots m$, *be independently*

*distributed as Wishart* $(n, p; \Sigma; S)$ *and multivariate normals* $(\mu_i, \Sigma)$ *and let*

$$x_i = \begin{pmatrix} x_{1\cdot i} \\ x_{2\cdot i} \end{pmatrix}\begin{matrix} q \\ p-q \end{matrix}, \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{matrix} q \\ p-q \end{matrix}, \quad \mu_i = \begin{pmatrix} \mu_{1\cdot i} \\ \mu_{2\cdot i} \end{pmatrix}\begin{matrix} q \\ p-q \end{matrix},$$

$$\Sigma = \begin{array}{c} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{matrix} q \\ p-q \end{matrix} \\ \begin{matrix} q & p-q \end{matrix} \end{array}, \quad \Delta^2_{q\cdot i} = \mu_{1\cdot i}' \Sigma_{11}^{-1} \mu_{1\cdot i} \text{ and } \Delta^2_{p\cdot i} = \mu_i' \Sigma^{-1} \mu_i .$$

*If*

$$\Delta^2_{p\cdot i} = \Delta^2_{q\cdot i}, \quad \text{then} \quad R_i = \frac{1 + x_{1\cdot i}' \left(S_{11} + \sum_{j=1}^{i-1} x_{1\cdot j} x_{1\cdot j}'\right)^{-1} x_{1\cdot i}}{1 + x_i' \left(S + \sum_{j=1}^{i-1} x_j x_j'\right)^{-1} x_i},$$

$i = 1, 2, \cdots m$ *and* $S_m = S + \sum_{j=1}^{m} x_j x_j'$ *are mutually independent and their distributions are given by const.* $R_i^{(n-p+i-2)/2}(1 - R_i)^{(p-q-2)/2}$ $i = 1, 2, \cdots m$ *and non-central Wishart* $(m + n, p, t; \Sigma, \mu; S_m)$ *with* $t$ *(*$t$ *is the rank of* $\mu = (\mu_1, \cdots \mu_m)$*) non-central parameters.*

The proof can be obtained in a similar manner as above.

*Note*: During the time of revision, the author has obtained the most general form of Theorem II which yields Theorem I as a corollary.

REFERENCES

[1] K. S. Rao, "On the mutual independence of a set of Hotelling $T^2$'s derivable from a sample of size $n$ from a $k$-variate normal population." *Proc. Int. Statistic, Conference India*, (1951), pp. 171–176.

[2] C. G. Khatri, "On the certain matrix, Jacobian and integral theorems useful in multivariate analysis" (unpublished).

[3] Deemer, W. L. and Ingram Olkin, "The Jacobians of certain matrix transformations useful in multivariate analysis," *Biometrika*, Vol. 38 (1951), pp. 345–367.

[4] K. V. Ramachandran and C. G. Khatri, "On the certain problems in multivariate analysis: I (Wishart distribution)," *Jour. M. S. University of Baroda*, Vol. VII (1958), pp. 79–82.

[5][1] S. N. Roy, "A report on some aspects of multivariate analysis," *North Caroline Institute of Statistics Mimeograph Series No.* 121.

[6][1] R. Gnanadesikan, "Contributions to multivariate analysis including univariate and multivariate variance components analysis and factor analysis," *North Carolina Institute of Statistics Mimeograph Series No.* 158.

[7][1] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1958.

[8] C. R. Rao, *Advanced Statistics in Biometric Research*, John Wiley and Sons, New York, 1951.

[1] I am thankful to the referee for suggesting these useful references.

# A NOTE ON THE STOCHASTIC INDEPENDENCE OF FUNCTIONS OF ORDER STATISTICS

By Gerald S. Rogers

*University of Arizona*

The theorem presented below appears to be useful in determining whether certain functions of order statistics are stochastically independent (briefly, independent). The following result has appeared in the literature in various forms, eg. [1]; the statement here is the particular form used in the proof of the second part of the theorem.

Lemma: *Let $s$ be a complete sufficient statistic for a family of probability density functions indexed by a parameter $\theta$. Let $t$ be any other statistic, not a function of $s$ alone. Then $s$ and $t$ are independent if and only if the distribution of $t$ does not depend on $\theta$.*

Theorem: *Let $x$ be a real random variable with distribution function $F(x) = \int_{-\infty}^{x} f(X)\,dX$, where $f(x)$ is a non-degenerate probability density function* (pdf). *Let $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ be the order statistics based on a random sample of size $n \geqq 2$ from this $x$ distribution. Let $z = z(x_1, \cdots, x_j)$ be a statistic based on the first $j < n$ order statistics only. Then the following two statements are equivalent:*

(a) *$z$ is independent of $x_k$ for some $k \geqq j$;*

(b) *$z$ is independent of the set $\{x_k : j \leqq k \leqq n\}$.*

Proof: Notation—let $g(A)[g(A|C)]$ denote the ordinary [conditional] pdf of $A$ [given $C$]. To show that (a) implies (b), first suppose that in (a), $k = j$. It follows directly from the definition of conditional pdf's that

$$g(x_1, \cdots, x_j|x_j) = g(x_1, \cdots, x_j|x_j, \cdots, x_n),$$

and hence that

$$g(z|x_j) = g(z|x_j, \cdots, x_n).$$

Under the hypothesis (a), $g(z|x_j) = g(z)$, and therefore, $g(z|x_j, \cdots, x_n) = g(z)$.[1] Thus, $z$ is independent of the set in (b).

Now suppose that in (a), $k > j$. Then, (as is readily shown by direct computation), in the conditional pdf $g(x_1, \cdots, x_{k-1}|x_k)$, $x_k$ may be considered as a "parameter" for which the conditional random variable $x_{k-1}$ given $x_k$, written $(x_{k-1}|x_k)$, is a "complete sufficient statistic." Under the hypothesis (a), $g(z|x_k) = g(z)$, so that the distribution of $z$ given $x_k$ actually does not depend upon the "parameter" $x_k$. Therefore, by the lemma, $(z|x_k)$ and $(x_{k-1}|x_k)$ are independent. In terms of the pdf's,

$$g(z, x_{k-1}|x_k) = g(z|x_k)g(x_{k-1}|x_k).$$

Since $g(z|x_k) = g(z)$, $g(z, x_{k-1}, x_k) = g(z)g(x_{k-1}, x_k)$, and hence, $g(z, x_{k-1}) = g(z)g(x_{k-1})$.

Received July 15, 1958; revised July 6, 1959.

[1] One referee pointed out that this is well known in the theory of Markov Chains.

This is a sufficient condition that $z$ and $x_{k-1}$ be independent. By repeated application of this process, it follows that $z$ is independent of $x_j$. But as shown above, this is also a sufficient condition that $z$ be independent of the set in (b). This completes the proof that (a) implies (b). That (b) implies (a) is evident.

The proof of an analogous theorem wherein $y = y(x_j, \cdots, x_n)$ is independent of the set $\{x_i: 1 \leqq i \leqq j\}$ is similar. Both theorems will also hold with respect to the sets $\{w_1, \cdots, w_j\}$ and $\{w_j, \cdots, w_n\}$ under a strictly monotone transformation $w = M(x)$. Moreover, since the theorems hold for the order statistics $x_1^* \leqq \cdots \leqq x_n^*$ obtained by sampling from a uniform distribution over $(0, 1)$, and since the transformation $x^* = F(x)$ is independent of the choice of points in intervals to which $F$ assigns zero probability [2], it follows that both theorems will hold under the weaker hypothesis that $F(x)$ is continuous. The truth or falsity of the theorems in the discrete or mixed cases remains an open question.

The author wishes to thank Professors A. T. Craig and R. V. Hogg under whose direction the theorem was evolved as part of a doctoral dissertation at the State University of Iowa.

## REFERENCES

[1] D. Basu, "On statistics independent of a complete sufficient statistic," *Sankhyā*, Vol. 15 (1955), pp. 377–380.

[2] H. Scheffé and J. W. Tukey, "Non-parametric estimation I. Validation of order statistics," *Ann. Math. Stat.* Vol. 16 (1945), pp. 187–192.

# CORRECTION NOTES

## CORRECTIONS TO "CONTINUED FRACTIONS FOR THE INCOMPLETE BETA FUNCTION"

By Leo A. Aroian

*Hughes Aircraft Company*

On page 218, lines 13 and 14, of this article (*Ann. Math. Stat.*, Vol. 12(1941), pp. 218–223), replace $b_{2s}$ and $b_{2s+1}$, by

$$b_{2s} = -\frac{(p+s-1)(q-s)}{(p+2s-2)(p+2s-1)}\frac{x}{1-x}$$

and

$$b_{2s+1} = \frac{s(p+q+s-1)}{(p+2s-1)(p+2s)}\frac{x}{1-x}$$

On page 220, line 2, replace $-1 < x < \infty$, by $-\infty < x < 1$.

On page 222 the statement is made that $I_{.8}(2.5, 1.5)$ could not be done by Müller's continued fraction. This is incorrect. Both continued fractions may be used for the range of $0 < x < 1$, and in this range both continued fractions equal $I_x(p, q)$.

On page 222 at the second line of Section 6, eliminate the words "due to the possible divergence of the series on which it is based." The rest of this paragraph is correct and both continued fractions may be used for $I_x(p, q)$.

## CORRECTIONS TO "SEQUENTIAL DECISION PROBLEMS FOR PROCESSES WITH CONTINUOUS TIME PARAMETER-TESTING HYPOTHESES"

By A. Dvoretzky, J. Kiefer, and J. Wolfowitz

The following corrections should be made on p. 259 of the above-titled paper (*Ann. Math. Stat.*, Vol. 24(1953), pp. 254–264): The mean occurrence time is $1/\lambda$, not $\lambda$, on line 4 of Section 4. In (4.2) the plus sign should be a minus sign.

## CORRECTIONS TO "DISTRIBUTION OF THE MAXIMUM OF THE ARITHMETIC MEAN OF CORRELATED RANDOM VARIABLES"

By John Gurland

*Iowa State University*

It has been called to my attention by P. R. Krishnaiah and M. M. Rao that the multivariate Gamma distribution with constant correlation between the

components is not unique. Although the particular case employed in the paper (*Ann. Math. Stat.*, Vol. 26 (1955), pp. 294–300) is stated unambiguously on the second page it is hoped that the following changes on the first page will help in avoiding any possible misinterpretation:

(i) Page 294. First sentence of summary; line 2. Insert the words "a particular case of a" before the words "multivariate analogue".

(ii) Page 294. Last sentence. Remove period at end of sentence and add the following: "and a special case of the multivariate analogue of the Pearson Type III distribution represented by (2)."

The following corrections are also kindly pointed out by Krishnaiah and Rao:

(iii) Page 294. Line 6 from bottom. Replace $\rho$ by $\rho^2$.

(iv) Page 295. Equation (2) is valid for $\lambda = n/2$ but not for all $\lambda > 0$. This does not affect the validity of the results obtained in the paper since the infinitely divisible distribution in (4) is valid for all $\lambda > 0$.

## CORRECTIONS TO "APPROXIMATION AND GRADUATION ACCORDING TO THE PRINCIPLE OF LEAST SQUARES BY ORTHOGONAL POLYNOMIALS"

By Charles Jordan

The following corrections should be made in the above-titled paper (*Ann. Math. Stat.*, Vol. 3(1932), pp. 257–357):

On page 335, instead of

$$\sum_{s=0}^{m+1} C_{ms} = 2m + 1 \quad \text{and} \quad \lim_{N\to\infty} C_{m0} = 2m + 1,$$

it should read

$$\sum_{s=0}^{m+1} | C_{ms} | = 2m + 1 \quad \text{and} \quad \lim_{N\to\infty} | C_{m0} | = 2m + 1.$$

On p. 356, $\binom{20}{10}$ should be 184756. In the original the last number is incorrect.

## CORRECTIONS TO "QUASI-RANGES IN SAMPLES FROM AN EXPONENTIAL POPULATION"

By Paul R. Rider

*Wright Air Development Center*

In the paper cited in the title (*Ann. Math. Stat.*, Vol. 30(1959), pp. 252–254), p. 253, fourth display, the exponent of the factor $e$ should be $-x_{r+1} - (r + 1)x_{n-r}$ instead of $-rx_{n-r}$. I thank Mr. George E. Bardwell for pointing this out.

The journal in reference [6] should be *Annals of the Institute of Statistical Mathematics* instead of *Ann. Math. Stat.*

## CORRECTIONS TO "ON BALANCING IN FACTORIAL EXPERIMENTS"

BY B. V. SHAH

*University of Bombay*

In the paper cited in the title (*Ann. Math. Stat.*, Vol. 29(1958), pp. 766–779), on p. 766, lines 23–26, the sentence should read as follows: "The set up assumed is that yield of a plot in the $j$th block having $i$th treatment is $\mu + \alpha_j + t_i + \epsilon_{ij}$, where $\mu$ is over all effect, $\alpha_j$ is the effect of the $j$th block, $t_i$ is the effect of the $i$th treatment and $\epsilon_{ij}$ is the experimental error."

On p. 776, line 2, change "any contrast" to "any normalised contrast".

On p. 777, in equations (7.7), (7.8) and (7.9), change "$(-1)^{q-1}$" to "$(-1)^{q-i}$".

I am indebted to a referee of a subsequent paper for pointing out these corrections.

## CORRECTION TO "A TABLE FOR COMPUTING TRIVARIATE NORMAL PROBABILITIES"

BY GEORGE P. STECK

*Sandia Corporation*

The following correction should be made to the paper of the above title (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 780–800):

Pages 790–799: replace "$m$" by "$h$" in the table headings.

## CORRECTIONS TO "A GENERALIZATION OF THE GLIVENKO-CANTELLI THEOREM"

BY HOWARD G. TUCKER

*University of California, Riverside*

The paper cited above (*Ann. Math. Stat.*, Vol. 30 (1959), pp. 828–830) contains several errors for which corrections are given below.

Inequality (4) should read

$$\sum_{j=1}^{k} F_n(X_{j-1,k})I_{A_j} \leqq F_n(x) \leqq \sum_{j=1}^{k} F_n(X_{jk} - 0)I_{A_j} \tag{4}$$

Inequality (6) should be replaced by

$$\begin{aligned} F(x \mid \mathfrak{Z}) - F_n(x) &\leqq \sum_{j=1}^{k} (F(X_{jk} - 0 \mid \mathfrak{Z}) - F_n(X_{j-1,k}))I_{A_j} \\ &= \sum_{j=1}^{k} (F(X_{jk} - 0 \mid \mathfrak{Z}) - F(X_{j-1,k} \mid \mathfrak{Z}))I_{A_j} \\ &\quad + \sum_{j=1}^{k} (F(X_{j-1,k} \mid \mathfrak{Z}) - F_n(X_{j-1,k}))I_{A_j} \\ &\leqq \max_{1 \leq j \leq k} \mid F_n(X_{jk}) - F(X_{jk} \mid \mathfrak{Z}) \mid + 1/k. \end{aligned} \tag{6}$$

Inequality (7) should be replaced by

$$F(x \mid \mathfrak{Z}) - F_n(x) \geqq -\max_{1 \leq j \leq k} \mid F_n(X_{jk} - 0) - F(X_{jk} - 0) \mid \mathfrak{Z}) \mid - 1/k. \tag{7}$$

Inequality (8) should be replaced by

$$\begin{aligned} \mid F_n(x) - F(x \mid \mathfrak{Z}) \mid \leqq 1/k + \max_{1 \leq j \leq k} \{\mid F_n(X_{jk} - 0) \\ - F(X_{jk} - 0 \mid \mathfrak{Z}) \mid, \mid F_n(X_{jk}) - F(X_{jk} \mid \mathfrak{Z}) \mid\}. \end{aligned} \tag{8}$$

Immediately after inequality (8) the following sentence should be added: In a way similar to the proof on the bottom of page 829 one may easily verify that $P[F_n(X_{jk} - 0) \underset{n}{\rightarrow} F(X_{jk} - 0 \mid \mathfrak{Z})] = 1$.

## CORRECTION TO "ON THE THEORY OF BAN ESTIMATES"[1]

By Robert A. Wijsman

*University of Illinois*

I am greatly indebted to Dr. Lucien LeCam for calling to my attention an error in the proof of Theorem 1 of the paper cited in the title (*Ann. Math. Stat.* Vol. 30 (1959), pp. 185–191). The transition from (12) to (13) is in general not justified. Worse, the theorem itself is false in general, as can be shown with a counter example. In order to remedy the situation, the assumptions have to be strengthened. This can be done either on the distributions of the $Z_n$, or on the estimator $\hat{\theta}$. As an example of the first, if the $Z_n$ have densities which (when normalized) converge a.e. to the limiting normal density, then the transition

[1] Work supported by the National Science Foundation.

from (12) to (13) is valid [9] and with that the proof of Theorem 1 is correct. However, this seems too strong an assumption to be of much practical value, since so many examples deal with discrete random variables. Turning now to assumptions on $\hat{\theta}$, we could require (4) to be true for all sequences $Z_n$ for which (1) holds. Taking then in particular $Z_n : N(\zeta(\theta), \Sigma(\theta)/n)$, the previous case (convergence of densities) applies, and the conclusion of Theorem 1 follows. A more attractive, even though slightly stronger, assumption on $\hat{\theta}$ is to require it to be differentiable in every point of $U$. This insures, of course, continuity in every point of $U$ but not continuity in a neighborhood of $U$, leave alone differentiability in a neighborhood of $U$ which would be the requirement for a regular (1) estimate. We are thus led to the following modification of Definition 2 and regular (2):

DEFINITION 3. $\hat{\theta}$ will be called regular (3) if (i) $\hat{\theta}(\zeta(\theta)) \equiv \theta$ identically in $\theta$[2]; (ii) $\hat{\theta}$ is differentiable in every point $\zeta(\theta)$ of $U$.

Let the matrix derivative of $\hat{\theta}$ in the point $\zeta(\theta)$ be denoted by $A(\theta)$. Theorem 1 now follows immediately by differentiation of (i) of Definition 3 (which is the same as equation (2)). A few remarks about $A(\theta)$ are in order. In the first place, the existence of this derivative in every point of $U$ implies (4) for every sequence $Z_n$ satisfying (1). Secondly, it is not necessary to require $A$ to be continuous in $\theta$. However, if $\hat{\theta}$ is constructed according to Theorem 2, then $A = (BV)^{-1}B$ (see eq. (6)) so that $A$ is continuous due to the continuity assumptions on $B$ and $V$. Under all circumstances, the $A$ corresponding to any BAN estimate is continuous since it is given by $A = (V'\Sigma^{-1}V)^{-1}V'\Sigma^{-1}$.

It is somewhat remarkable that Theorem 2 remains true if, in the conclusion, regular (2) is replaced by the stronger regular (3). The surprise is that $\hat{\theta}$ turns out to be differentiable in each point of $U$, even though no differentiability assumptions are made on $B$. Therefore, a proof of Theorem 2, with regular (2) replaced by regular (3), seems to be in order. Before doing this, it may be of interest to point out that Ferguson's estimates [5] are also differentiable in each point of $U$ since they are generated by (5) with $B(z, \theta)$ satisfying even stronger assumptions than in Theorem 2. Comparing now the various kinds of regular estimates, we have that regular (1) estimates are continuously differentiable in a neighborhood of $U$, Ferguson's estimates are continuous in a neighborhood of $U$ and differentiable in every point of $U$, while regular (3) estimates are differentiable in every point of $U$.

PROOF OF THEOREM 2, with regular (2) replaced by regular (3). It suffices to show that in each point of $U$ there is a neighborhood possessing the properties ascribed to the neighborhood $N$ in the conclusion of Theorem 2. Then $N$ can be taken as the union of the individual neighborhoods. Consider any point of $U$. We may take this as the origin of the coordinate system in Z. For the purpose of the proof we may make the same transformations as in Section 4 (observe

[2] The assumption (i) of Definition 3 is the same as equation (2). Instead, we could have made the same assumption as in Definition 1 (i). The two assumptions are equivalent since $\hat{\theta}$ is supposed to be continuous in each point of $U$.

that $\zeta^{-1}$ is differentiable due to Assumption 2 (iii) and (iv)). We may suppose then that $U$ is a linear subspace of $Z$, spanned by the first $m$ coordinate axes, and that $\zeta$ is the identity function from $U$ to $U$. Thus we have identified $U$ with the parameter space $\Omega$. A point $u$ of $U$ has its last $k - m$ components equal to 0; the $m$-vector formed by its first $m$ components will be written $\theta$. Let $I_{km}$ be a $k \times m$ matrix whose elements are 1 on the "main diagonal" and 0 otherwise. We can write then $u = I_{km}\theta$. The transformations which we have employed replace in (5) $\zeta(\theta)$ by $I_{km}\theta$, and $B(z, \theta)$ by some other matrix, which, however, we shall again denote by $B(z, \theta)$. The matrix $V(0)$ is replaced by $I_{km}$. Put $B(z, \theta)\, I_{km} = C(z, \theta)$, then by assumption $C(0, 0)$ is non-singular. Furthermore, $C$ is continuous in $(z, \theta)$ at $(0, 0)$. Put $C^{-1}B = D$, then $D(z, \theta)$ exists in a neighborhood of $(0, 0)$, is continuous in $(z, \theta)$ at $(0, 0)$ and is continuous in $\theta$ for each fixed $z$. Let $S_1 \times S_2$ be such a neighborhood, where $S_1$ is a solid $k$-sphere about $z = 0$ and $S_2$ a solid $m$-sphere about $\theta = 0$. In addition, we may choose the radii $r_1$ and $r_2$ of $S_1$ and $S_2$ so that for $(z, \theta) \in S_1 \times S_2$ we have $\| D(z, \theta) \| \leqq r_2/r_1$. We now write (5) as

$$\theta = D(z, \theta)z. \tag{25}$$

For each $z \in S_1$, the right hand side of (25) is a continuous transformation of $S_2$ into itself. According to the Brouwer fixed point theorem [7] there is a fixed point of the transformation, therefore a solution $\hat{\theta}(z)$ to (25). Write

$$\hat{\theta}(z) = D(z, \hat{\theta}(z))z. \tag{26}$$

For $z \in S_1$, $\| D(z, \hat{\theta}(z)) \|$ is bounded, so $\hat{\theta}(z) \to 0$ as $z \to 0$. Hence $\hat{\theta}$ is continuous at 0. From this we have $D(z, \hat{\theta}(z)) \to D(0, 0)$ as $z \to 0$, and from (26) it follows then that $\hat{\theta}$ is differentiable at $z = 0$, with matrix derivative $D(0, 0)$. This proves that on $S_1$ $\hat{\theta}$ is regular (3). In the original coordinate system the matrix $D(0, 0)$ takes the form $(BV)^{-1}B$, evaluated at some point $(\zeta(\theta), \theta)$. This leads immediately to (6). The last assertion in the conclusion of Theorem 2 is proved in [3].

## REFERENCE

[9] HENRY SCHEFFÉ, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 434–438.

## ABSTRACTS OF PAPERS

*(Abstracts of papers not presented at any meeting of the Institute)*

**6. On a $\chi^2$-Test with Cells Determined by Order Statistics.** HERMANN WITTING, University of Freiburg. (By title)

Let $X_1, \cdots, X_n$ be a sample of a continuous one-dimensional probability distribution $Q(A)$; let $X_{n,1}, \cdots, X_{n,k-1}$ be order statistics for given ranks $r_{n,j}$ with $p_{n,j} = (r_{n,j} - r_{n,j-1})/(n+1) = p_j + o(1/\sqrt{n})$. Let $S_{n,j} = \{x: X_{n,j-1} < x \leqq X_{n,j}\}$. For testing the hypothesis that $Q(A)$ belongs to an $s$-parametric class of probability distributions $P(A, \theta)$ the test statistic $T_n = \sum_{j=1}^{k} n(P(S_{n,j}, \hat{\theta}_n) - p_{n,j})^2/p_{n,j}$ is used, where $\hat{\theta}_n$ is the minimum-$\chi^2$-estimate. Then if $Q(A) = P(A, \theta_0)$ or $Q(A) = P(A, \theta_0) - q(A)/\sqrt{n}$, respectively, under certain regularity conditions $T_n$ is asymptotically distributed as $\chi^2$ with $(k - s - 1)$ degrees of freedom (and noncentrality parameter $\sum_{j=1}^{k} q_j^2/p_j$, $q_j = p\text{-lim}\, q(S_{n,j})$. Using $(k-1)$ continuous functions $\varphi_1(x), \cdots, \varphi_{k-1}(x)$, defining $\varphi_j(X_{n,j})$ successively by ordering the values $\varphi_j(X_i)$ and defining $S_{n,j} = \{x: \varphi_l(x) > \varphi_l(X_{n,l}), l = 1, \cdots, j-1; \varphi_j(x) \leqq \varphi_j(X_{n,j})\}$, the same limiting behaviour of $T_n$ holds for probability distributions in a metric space. The proof is based on the fact that the $Q(S_{n,j})$ are jointly $B$-distributed (cf. J. W. Tukey, *Ann. Math. Stat.* 18(1947)529). Therefore $\sqrt{n}(Q(S_{n,j}) - p_{n,j})$ are asymptotically $N(0, C)$ where $C$ is of rank $(k-1)$ and coincides with the covariance matrix of the multinomial distribution, underlying the corresponding classical $\chi^2$-test with the cells $S_j = p\text{-lim}\, S_{n,j}$. While having the same power, this modified $\chi^2$-procedure has certain advantages over the classical $\chi^2$-test.

**7. A Generalized Pitman Efficiency for Nonparametric Tests.** HERMANN WITTING, University of Freiburg. (By title)

Asymptotic expressions up to terms of order $n^{-2}$ are given for the efficiency of the Wilcoxon two-sample test relative to the $\bar{x}$- and $t$-tests for nearby alternatives. The first term is the well-known Pitman efficiency; the remaining terms are corrections for finite sample sizes. Efficiency values are given for finite sample sizes in the case of normal and rectangular distributions and comparisons with the exact values are made. In general the Wilcoxon test is shown to be nearly as good locally for moderate sample sizes as it is known to be asymptotically. A similar analysis is performed for the single-sample sign test.

*(Abstracts of papers to be presented at the Washington, D. C., Annual Meeting of the Institute, December 27–30, 1959. Additional abstracts will appear in the March, 1960 issue.)*

**1. Some Nonparametric Problems: I.** V. P. BHAPKAR, University of North Carolina and University of Poona. (By title)

Mood and Brown have considered a nonparametric test for the equality of row effects in the two-way classification with one observation per cell or the same number of observations per cell. In this paper, first their test has been extended to cover incomplete block situations. For the BIBD in the usual terminology, if $m_i$ denotes the number of observations, for the $i$th 'treatment', that exceed the respective 'block'-medians, then to test the equality of 'treatment'-effects we have $(k^2(k-1))/(a(k-a)\lambda v) \sum_{i=1}^{v} (m_i - (ra/k))^2$ asymptotically distributed as a $\chi^2$ with $v - 1$ d.f. for large $r$, where $a$ is $k/2$ if $k$ is even and $(k-1)/2$ otherwise. The $\chi^2$ statistic appropriate for PBIBD is also given.

Next, Hoeffding's theorem on $U$-statistics extended by Lehmann to the case of two samples, has been extended to the case of $c$ samples. This is then applied to derive a new

test for the problem of $c$ samples. The test criterion is in terms of the number of $c$-plets that can be formed by choosing one observation from each sample such that the observation from the $k$th sample is the least $(k = 1, 2, \cdots, c)$.

2. **Some Nonparametric Problems: II.** V. P. Bhapkar, University of North Carolina and University of Poona. (By title)

Mood and Brown have considered some simple nonparametric regression problems. In this paper, their methods are extended to discuss some additional regression problems. Next some bivariate analysis of variance problems are considered. The step-down procedure is used to reduce the problem to one involving conditional univariate distributions, the other variate being regarded as a concomitant random variate. The regression methods developed earlier are used here in these bivariate problems. The method seems to be perfectly general and could be extended to the general multivariate situation.

3. **On the Foundations of the Theory of Testing Hypotheses** (Preliminary report). Allan Birnbaum, New York University.

For testing between simple hypotheses $H_i$, $i = 1, 2$, an experiment is called *simple* if it is equivalent, in the sense of the theory of comparison of experiments, to one observation on $X$, where $\text{Prob}[X = 1 \mid H_i] = p_i$, $\text{Prob}[X = 0 \mid H_i] = q_i = 1 - p_i$, $i = 1, 2$, with $p_i$'s known, $0 \leqq p_1 \leqq p_2 \leqq 1$. If various experiments are possible for a given testing problem, and if one of these is selected by use of a definite random device unrelated to the hypotheses, the over-all procedure is called a mixture of experiments. It is proved that under minor restrictions every experiment is equivalent to a mixture of simple experiments called its *components*. The possible decompositions into components are characterized and shown to be not essentially unique, except for simple experiments, whose components are equivalent to the given experiment. It follows that customary interpretations of error-probabilities of a test, as indicators of strength of evidence provided by a test outcome, require critical and constructive revision which leads to a modified Neyman-Pearson theory in which the likelihood function holds a central position as a consistently interpretable primitive indicator of evidence relevant to hypotheses. Wald's sequential test is given an elementary justification on these terms as a technique for informative inference.

4. **Unbiased Sequential Estimation for Certain Two Parameter Problems** (Preliminary report). B. Brainerd, University of Western Ontario, I. Chorneyko, University of Alberta, and T. V. Narayana, University of Alberta.

The probability of a coin falling head is $p_1(0 < p_1 < 1)$, if in the previous trial the outcome was tail and $p_2(0 < p_2 < 1)$, if in the previous trial the outcome was head. At the first trial the probability of a head is $p_1$. Using a technique devised by one of the authors, sufficient partitions are obtained for a wide class of simple closed regions. The results of M. H. DeGroot (*Ann. Math. Stat.*, Vol. 30 pp. 80–102) are shown to generalize, with the proper modifications, to the two parameter case. Estimable functions are explicitly given, and completeness of sampling plans proved for various regions. An analogue of the necessary and sufficient conditions of Lehmann and Stein for simple closed regions is being studied.

5. **Mathematical Models for Ranking from Paired Comparisons.** H. D. Brunk, University of Missouri. (By title)

Several models are discussed in each of two categories: (I) Each possible ranking of items is assumed to have a "utility" (for some segment of the community) which depends on the expected scores of the items in paired comparisons. Special instances are models in which

"worth" of an item is *defined* in terms of its expected scores in comparisons with other items. (II) Each item is *assumed to have an intrinsic worth*; these intrinsic worths determine the expected scores.

The concept of "regularity" is introduced. Let the expected scores of Item A be at least as large as those of Item B. A utility is *regular* if under these conditions every ranking in which Item A precedes Item B has at least as great utility as one in which they are interchanged. This concept specializes to rankings based on worths. A necessary and sufficient condition is given in order that a linear utility may be regular.

In the second category a "minimum assumption" model is introduced and discussed. Let $e(u, v)$ denote the expected score of an item of worth $u$ when compared with one of worth $v$. The assumption is: $e(u, v)$ is non-decreasing in $u$, non-increasing in $v$.

6. **Asymptotically Optimal Stopping Rules in Sequential Analysis** (Preliminary Report). HERMAN CHERNOFF, Stanford University.

It is desired to decide sequentially whether the mean $\mu$ of a normal distribution with known variance is positive or negative. Suppose that an a priori distribution is given for $\mu$ which has positive density at $\mu = 0$. Suppose also that the loss due to coming to the wrong conclusion is given by $r(\mu) = k \mid \mu \mid + 0(1)$, as $\mu \to 0$. Finally suppose that the cost of sampling $c \to 0$. For the optimal sequential procedure the main contribution to the Bayes risk is given by those values of $\mu$ which are of the order of magnitude of $c^{1/3}$.

The optimal stopping rule is approximated by the solution of the analogous continuous problem involving a Wiener process. This problem in turn is reduced to the solution of a free boundary problem involving the heat equation. A method of constructing this boundary is proposed.

7. **Cross-Compounded Distributions.** RICHARD A. EPSTEIN AND LLOYD R. WELCH, Jet Propulsion Laboratory, California Institute of Technology. (Introduced by L. A. Zadeh)

It is known that the generating function of the compound Poisson distribution has the property that the Poisson variable can be expressed as the sum of two or more independent variables. A particular method of "cross-compounding" two distributions is suggested; the same property obtains. In theory, any two distributions can be cross-compounded to produce a third, unique distribution. However, frequently the mathematics become overly involved so that it is necessary to select the distributions with discretion. Examples are given wherein the negative binomial distribution is cross-compounded with the Exponential distribution and with the Poisson distribution. Other combinations are also suggested which might lend themselves to cross-compounding.

8. **Examples of Two Independent Separable Processes Whose Sum Is Not Separable.** T. FERGUSON, Princeton University.

Two examples are given of two independent stochastic processes, $X_t$ and $Y_t$, both of which are separable in the sense defined in Doob's book, and yet whose sum $Z_t = X_t + Y_t$ is not a separable process. All processes considered are measurable. In the first example, $X_t$ is a constant (i.e. non-random) function, while in the second example, $X_t$ and $Y_t$ are identically distributed.

9. **On the Exactness of the Missing Plot Procedure in a Randomized Block Design.** J. L. FOLKS AND D. L. WEST, Texas Instruments Incorporated.

The randomized block design is said to be an unbiased design in that it allows unbiased estimates of treatment differences, an unbiased estimate of the error variance and an un-

biased test of treatment differences. This claim can be justified by assuming that the observations are generated by the model $y_{ij} = \mu + b_i + t_j + e_{ij}$ where $e_{ij} \sim N(0, \sigma^2)$. It can also be justified by considering the population of conceptual yields arising from all possible randomizations. In the case of a missing plot, the exact procedure described by Yates gives unbiased estimates of treatment differences, an unbiased estimate of the error variance, and an unbiased test of treatment differences. However, it is based only on the normal model, not upon randomization theory. The authors examine the missing plot procedure from the standpoint of randomization theory. The finite population of conceptual yields is examined where (1) the same block-treatment combination is always missing, and where (2) the same block-plot combination is always missing. In both cases, unbiased estimates of treatment differences are given by the usual estimates. The estimate of error variance and the test for treatment differences are unbiased in (1) under a restriction slightly weaker than homogeneity but appear not to be in (2) for any reasonable restriction.

### 10. First Emptiness of Two Dams in Parallel. Joseph M. Gani, Columbia University.

The paper considers the probabilities of first emptiness of two dams in parallel, both subject to steady releases at a constant unit rate, and fed by discrete Poisson inputs of unit size which are directed to the dam with lesser content. The problem is shown to be equivalent to that of the single dam fed alternately by the two ordered inputs $0 \leqq \alpha, \beta \leqq 1 (\alpha + \beta = 1)$; starting with an initial content $z$, the probabilities of first emptiness of the process beginning with an input $\alpha$, at the times

$$T = z + [(n + 1)/2]\alpha + [n/2]\beta \ (n = 0, 1, 2, \cdots)$$

are given by $g_\alpha(z, T) = e^{-\lambda z}$ if $n = 0$, and $g_\alpha(z, T) = e^{-\lambda z}\{ \sum_{j=1}^{[(n+1)/2]} g_\beta(j\alpha + j\beta - \beta, [(n + 1)/2]\alpha + [n/2]\beta)(\lambda z)^{2j-1}/(2j - 1)! + \sum_{j=1}^{[n/2]} g_\alpha(j\alpha + j\beta, [(n + 1)/2]\alpha + [n/2]\beta) (\lambda z)^{2j}/(2j)!\}$ if $n = 1, 2, \cdots$, where $g_\beta(z, z + [(k + 1)/2]\beta + [k/2]\alpha) (k = 0, 1, 2, \cdots)$, the analogous probability beginning with an input $\beta$, is given by an interchange of $\beta$ for $\alpha$ in the previous equation. These probabilities may be evaluated recursively. A more convenient method is found by reducing the process to an associated occupancy problem, when the probabilities can be obtained by a rapid computational procedure. Generating functions of the probabilities are derived, and the paper concludes with a general formulation of the dam problem when the times of arrival for two ordered non-negative inputs of random size form a Poisson process.

### 11. Stochastic Approximation and "Minimax" Problems. L. A. Gardner, Jr., MIT Lincoln Laboratory. (By title)

With the exception of the Robbins-Monro and Kiefer-Wolfowitz processes, the technique of stochastic approximation does not appear to have found a range of application consistent with the generality of its formulation (for exposition see C. Derman, "Stochastic approximation", *Ann. Math. Stat.*, Vol. 27 (1956), pp. 879–886). In this paper we consider such an iterative scheme designed to estimate the minimum of a curve which is not a regression function but the a.s. supremum of an observable random variable depending upon a parameter. The range of the parameter is a known finite interval, and the possibility of the solution being a boundary point is admitted. "Deterministic" conditions of the usual kind are imposed. The procedure is formally a truncated Kiefer-Wolfowitz process with the estimate of slope calculated from observed largest values in samples whose size tend to infinity as the iteration proceeds. Convergence with probability one is insured if this number increases sufficiently fast, or equivalently the differencing interval decreases suffi-

ciently slow, relative to a measure of the amount of probability in left neighborhoods of the function to be minimized. Although it is easy to argue the existence of such a measure, it cannot be assumed that anything is known concerning its (finite) value. This difficulty is resolved by having the differencing interval to be used for obtaining the next iterant depend in an appropriate way on a sample of largest values at the present. Estimates of convergence rates are made and optimum values obtained for certain constants of the process. Examples show the applicability of the theory to diverse problems.

12. **Some Asymptotic Results for a Coverage Problem.** MAX HALPERIN, Knolls Atomic Power Laboratory. (By title)

Let $\Delta_1, \Delta_2, \cdots, \Delta_n$ be a random sample from a population with probability density $p(\Delta)$, $0 \leqq \Delta \leqq \Delta_M$, $\Delta_M$ finite. The set of line segments corresponding to the $\Delta_i$ are cast on the interval $(0, L)$, $L \geqq n\Delta_M$, in such a way that every admissible configuration of the segments is equally likely. A configuration is admissible if (a) there is no overlapping of segments with each other. (b) there is no overlapping of segments with 0 or $L$. Now suppose a line of length $\lambda$ is cast at random in the interval $(0, L), \lambda < L$; i.e. $y$, the coordinate of the midpoint of the line of length $\lambda$ is distributed uniformly on $(\lambda/2; L - (\lambda/2)$. We define fractional converage, $F$, as the fraction of the line of length $\lambda$ which is covered by the segments of length $\Delta_1, \Delta_2, \cdots, \Delta_n$ and consider the probability distribution of $F$ as $n, L \to \infty$ and $n\mu/L \to V$ where $\mu = E\Delta$ and $0 < V < 1$. It is shown that $\Pr\{F = 0\} = (1 - V) \exp - (V\lambda/(1 - V)\mu)$, $\Pr\{F = 1\} = V/\mu \int_\lambda^{\Delta_M} (y - \lambda)p(y)\,dy$, if $\lambda > \Delta_M$ and in zero if $\lambda \geqq \Delta_M$; for $0 < F < 1$, there are further (continuous) contributions to the cumulative probability which unfortunately are critically dependent upon the nature of $p(\Delta)$. One can show that $EF = V$ independently of the specific nature of $p(\Delta)$ for $\lambda > \Delta_M$ but the variance is a complex function of $p(\Delta)$ which is not simply expressible even for specific $p(\Delta)$. It can be shown that for large $\lambda$, $F$ is normally distributed with mean $V$ and variance $\mu\, V(1 - V)^2[1 + \sigma^2/\mu^2]/\lambda$ where $\sigma^2 = E\Delta^2 - \mu^2$.

The above work was motivated by the need for a plausible graduation function to fit the distribution of Boron Carbide intercepted by neutron paths (Boron Carbide is used to control reactor power output). Although the above assumptions are quite naive relative to the actual complexity of the problem, preliminary experimental data suggests that use of the results to match a graduation function to two moments may adequately describe observed frequency distributions.

13. **Polya Type Distributions of Convolutions.** SAMUEL KARLIN, Stanford University, AND FRANK PROSCHAN, Sylvania Electric Products, Inc., Mt. View, California.

This paper obtains several useful new theorems concerning successive convolutions of Polya frequency densities, such as: If $f_1, f_2, \cdots$ are density of non-negative random variables with each $f_i$ a Polya frequency density of order $k$, then $g(n, x) = f_1^* f_2^* \cdots {}^*f_n(x)$ (the $n$-fold convolution) is Polya type of order $k$ in the variables $n$ and $x$, where $n$ ranges over the positive integers and $x$ traverses the positive real line. More generally, the following theorem is derived: Let $f_1, f_2, \cdots$ be a sequence of Polya frequency densities of order $k$ for corresponding general real valued (not necessarily positive) random variables $X_1$, $X_2, \cdots$. Then $h(n, x) = P[\sum_{i=1}^n X_i \geqq x;\ \sum_{i=1}^j X_i < x, j = 1, 2, \cdots, n-1]$ is totally positive of order $k$ in $n$ and $x$, $n$ ranging over the positive integers and $x$ over the positive axis. Applications of these theorems are given in inventory theory, probability, and mathematics.

**14. A New Proof of the Continuity Theorem of Probability Theory.** EMANUEL PARZEN, Stanford University. (By title)

The continuity theorem states that if a sequence of characteristic functions $\varphi_n(t)$ converge to a characteristic function $\varphi(t)$ at each real $t$, then the corresponding distribution functions converge, $F_n(x) \to F(x)$ at all continuity points $x$ of $F$. The presently known proofs of this theorem are not constructive, but rather involve compactness arguments. This paper gives a new constructive proof of the continuity theorem, based on the observation that $\int_{-\infty}^{\infty} g(x)\, dF_n(x) \to \int_{-\infty}^{\infty} g(x)\, dF(x)$ for any bounded continuous function $g$ with integrable Fourier transform. Details of the proof are given in Chapter 10 of my book *Modern Probability Theory and its Applications*, John Wiley, New York, 1960.

**15. A New Inversion Formula.** EMANUEL PARZEN, Stanford University. (By title)

Let $g$ be a bounded integrable Borel function of a real variable which possesses right and left hand limits at every real $x$. Let $g^*(x) = \{g(x + 0) + g(x - 0)\}/2$. Let

$$\gamma(u) = (1/2\pi) \int_{-\infty}^{\infty} e^{-iux} g(x)\, dx.$$

Then for any distribution function $F$ (with corresponding characteristic function $\varphi$) $\int_{-\infty}^{\infty} g^*(x)\, dF(x) = \lim_{U\to\infty} \int_{-U}^{U} (1 - (| u |/U))\gamma(u)\varphi(u)\, du$. The proof is given in Chapter 9 of my book *Modern Probability Theory and its Applications*, John Wiley, New York, 1960.

**16. A Law of Large Numbers for Dependent Random Variables.** EMANUEL PARZEN, Stanford University. (By title)

Let $X_1, X_2, \cdots$ be random variables with zero means and uniformly bounded variances. Let $Z_n = (X_1 + \cdots + X_n)/n$. Let $C_n = E[X_n Z_n]$. *Quadratic Mean Law of Large Numbers.* $Z_n \to 0$ in mean square as $n \to \infty$ if and only if $C_n \to 0$ as $n \to \infty$. *Strong Law of Large Numbers.* $Z_n \to 0$ with probability one if $C_n = 0(n^q)$ for some positive $q$. These results generalize some of the known laws of large numbers for orthogonal and stationary sequences of random variables. The proof is based on the identity $n^2E[Z_n^2] + \sum_{k=1}^{n} E[X_k^2] = 2 \sum_{k=1}^{n} kC_k$. Details are given in Chapter 10 of my book *Modern Probability Theory and its Applications*, John Wiley, New York, 1960.

**17. Inference in Stochastic Processes I: Testing Composite Hypotheses** (Preliminary report). M. M. RAO, Carnegie Institute of Technology.

Let $\{x(t), t \varepsilon T\}$ be a (real) stochastic process where $T$ is a linear Borel set. For any $n$, let $t_1 < t_2 < \cdots < t_n$ be in $D$, a dense subset of $T$, and $f_{t_1,\cdots,t_n}(x_{t_1}, \cdots, x_{t_n}; \theta)$, or $f_n(x, \theta)$ say, be the finite dimensional density function (w.r.t. Lebesgue meas.), of the process, which depends on $\theta = (\theta_1, \cdots, \theta_k)$, $k$ being independent of $n$. Suppose the testing problem consists of the hypotheses $H_0 : \theta \varepsilon \omega_a$ vs. $H_1 : \theta \varepsilon \omega_r$ (based on one realization), where $\omega_a$ and $\omega_r$ are closed disjoint subsets of the (real) Euclichian $k$-space. Assume the following conditions on the densities: (a) for all $n$, the carriers of $f_n(x, \theta)$ remain invariant for all $\theta$ in $\Omega = \omega_a + \omega_r$, and $f_n$ are Baire densities, (b) if $\theta_1$, and $\theta_2$ in $\Omega$ are distinct, then $f_n(x, \theta_1) \neq f_n(x, \theta_2)$ a.e., and (c) if $\xi(\theta)$ is any distribution function (d.f.) on $\Omega$ which assigns positive probability to both $\omega_a$ and $\omega_r$, then $(f_{n+1}(x, \theta) \int_\omega f_n(x, \theta)\, d\xi(\theta) - f_n(x, \theta) \int_\omega f_{n+1}(x, \theta)\, d\xi(\theta))$, for any $\theta$ in $\omega(= \omega_a$ or $\omega_r)$, is either non-negative or non-positive for all $n$.

*Theorem: If* $\{x(t), t \varepsilon T\}$ *is a real separable stochastic process without fixed points of discontinuity and with the finite dimensional density functions* $f_n(x, \theta)$ *satisfying the conditions* (a)–(c), *then, for a sufficiently large number of observations on the process at* $t_i$ *of* $D$, *there exists an essentially unique Bayes solution, relative to an a priori distribution* $\xi(\theta)$ *on* $\Omega$, *satisfying* (c), *for testing the composite hypotheses* $H_0$ *against* $H_1$. Instead, $\theta$, being a vector of $k$ components, may depend on $t$ (or $k$ or $n$). Then, if the condition (c) is suitably modified, an analogous result obtains. Some applications are considered.

### 18. **Testing of Hypotheses on Categorical Data.** S. N. Roy, University of North Carolina, and V. P. Bhapkar, University of Poona.

In an earlier paper, we have posed hypotheses, which might be considered to be generalizations, appropriate to the categorical data (structured or unstructured), of the usual hypotheses in classical 'normal' univariate and multivariate analysis of variance and in analysis of various kinds of 'normal' association. The large sample tests for some such hypotheses have been offered earlier and for most of the rest are offered here. The theorem on minimum $\chi_1^2$ is proved along Cramér's lines and an independent justification for Neyman's 'linearization' technique is given. It is also shown that for linear hypotheses the minimum $\chi_1^2$ is exactly the same expression as the minimum sum of squares obtained by the "general least squares" approach to a model involving some asymptotically normal variables.

### 19. **On Tests of Certain Types of Hypotheses Involving the Dispersion Matrices of Two or More Multivariate Normal Distributions and the Associated Confidence Bounds.** S. N. Roy, University of North Carolina, and R. Gnanadesikan, Bell Telephone Laboratories.

For $N\left(\underset{p\times 1}{\xi_i}, \underset{p\times p}{\Sigma_i}\right) (i = 1, 2)$, one of the authors derived several years ago, on a certain principle, a test for $H_0 : \Sigma_1 = \Sigma_2$ against $H: \Sigma_i \neq \Sigma_2$, with an acceptance region $\mu_1 \leqq$ all ch $(S_1S_2^{-1}) \leqq \mu_2$, where $S_1$ and $S_2$ are the sample dispersion matrices, and also the associated confidence bounds. In this paper the same principle is used to derive tests for $H_0 : \Sigma_1 = \Sigma_2$ against the respective alternatives (i) $H$: all ch $(\Sigma_1\Sigma_2^{-1}) > 1$, (ii) $H$: all ch $(\Sigma_1\Sigma_2^{-1}) < 1$, (iii) $H$: (i) $\cup$ (ii), (iv) $H$: at least one ch $(\Sigma_1\Sigma_2^{-1}) > 1$ and (v) $H$: at least one ch $(\Sigma_1\Sigma_2^{-1}) < 1$. The associated confidence bounds are also obtained and interpreted, and finally, a partial generalization of these results are made to the case of $k$ populations, with regard to both testing of hypotheses and confidence bounds.

### 20. **On the Monotonic Character of the Power Functions of Two Multivariate Tests.** S. N. Roy and W. F. Mikhail, University of North Carolina.

The power function of the largest root test of normal multivariate linear hypothesis on means or of independence between two sets of variates involves, in each case, aside from the degrees of freedom, certain non-negative, non-centrality parameters. This paper supplies a relatively simple and compact proof that the power function monotonically increases as each parameter, separately, increases—a result that was conjectured and proved (but not published) by one of the authors several years ago by a very lengthy and laborious method. It is believed that, with suitable and slight modifications, the method used here should prove useful in proving or disproving similar results in a wide variety of problems in testing of hypotheses involving multivariate normal distributions.

**21. Confidence Bounds for an Integral Function of an Estimate with Applications to Reliability Theory.** SAM C. SAUNDERS, Boeing Scientific Research Laboratories.

Let $X$ and $Y$ be independent random variables with distributions $F \varepsilon \mathfrak{F}$ and $G \varepsilon \mathfrak{G}$, respectively, where $\mathfrak{F}$ and $\mathfrak{G}$ are subsets of the class of continuous distributions on given positive sample spaces $\mathfrak{X}$ and $\mathfrak{Y}$. Let $\omega$ be a homeomorphism from $\mathfrak{Y}$ into $\mathfrak{X}$ and define $H(\omega) = \int F(\omega)\, dG$. From samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ we form $\tilde{F}_n$ and $\tilde{G}_m$, estimates of $F$ and $G$, respectively, and define $\tilde{H}$, the empirical estimate of $H$, by $\tilde{H}(\omega) = \int \tilde{F}_n(\omega)\, d\tilde{G}_m$ for $\omega \varepsilon \Omega$, a class of homeomorphisms linearly ordered by $H$.

We are interested in problems associated with this phenomenological interpretation. For some device: let $\omega(Y)$ be the taxation on life under usage $\omega$ and let $X$ be the capacity for endurance. Then $H(\omega) = P[X < \omega(Y)]$ is the unreliability and $\tilde{H}(\omega)$ is an estimate of this unreliability. Using $\tilde{H}$ to determine a maximum usage $\tilde{\omega}$, what is the probability the unreliability $H(\tilde{\omega})$ is too large? We define $\tilde{\omega}$ so that $H(\tilde{\omega})$ is distribution-free re $\mathfrak{F} \times \mathfrak{G}$ or obtain a stochastic bound majorizing $H(\tilde{\omega})$ for each $(F, G) \varepsilon \mathfrak{F} \times \mathfrak{G}$ under the assumption $\tilde{F}(F^{-1})$ is distribution-free re $\mathfrak{F}$ and similarly for $\tilde{G}$, $\mathfrak{G}$. This provides an answer in one important application and the theory is developed so that many such reliability problems can be treated.

**22. A Rank Sum Test for Comparing all Pairs of Treatments.** ROBERT G. D. STEEL, Cornell University.

Consider a permutation of $n_1X_1$'s, $\cdots$, $n_kX_k$'s with $n_1 \leqq \cdots \leqq n_k$ arising from ordering, from smallest to largest, observations on $k$ treatments. Assign ranks $1, \cdots, n_i + n_j$ to the observations on all possible pairs of treatments and sum the ranks assigned to the observations on the treatment with lower subscript. This gives a test criterion denoted by $(T_{12}, \cdots, T_{1k}, T_{23}, \cdots, T_{k-1,k})$. A recursion formula is developed for computing probabilities and is used to show, by induction, that $\mu(T_{ij}) = n_i(n_i + n_j + 1)/2$, $\sigma^2(T_{ij}) = n_in_j(n_i + n_j + 1)/12$, $\sigma(T_{hi}T_{hj}) = n_hn_in_j/12 = \sigma(T_{hj}T_{ij})$, $\sigma(T_{hi}T_{ij}) = -n_hn_in_j/12$ and $\sigma(T_{gh}T_{ij}) = 0$. From the distribution of $(T_{12}, \cdots, T_{k-1,k})$, the distribution of $\min\{T_{ij}\}$ can be obtained. Several such distributions are computed for a common value of $n$. These provide critical values for a non-parametric multiple comparison rank sum test.

**23. Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient.** JOHN S. WHITE, Aero Division, Minneapolis Honeywell Regulator Co.

Following the procedure used by W. J. Dixon (*Ann. Math. Stat.*, 1944, pp. 119–144) series expansions are obtained for the first two moments of $\hat{\alpha} = \sum x_t x_{t-1}/\sum x_{t-1}^2$ where $(x_t)$ is a first order auto-regressive Gaussian process with parameter $\alpha$. The series expansions are carried out to terms of order $T^{-3}$ and $\alpha^4$ thus extending the asymptotic results of several authors.

The results are obtained for both the stationary and fixed initial variate case.

## NEWS AND NOTICES

*Readers are invited to submit to the Secretary of the Institute news items of interest*

### Personal Items

Professor Robert Bechhofer, having spent his sabbatic leave at Stanford University, has returned to the Department of Industrial and Engineering Administration, Sibley School of Mechanical Engineering, Cornell University. While at Stanford, Dr. Bechhofer held an appointment as Visiting Professor of Statistics in the Applied Mathematics and Statistics Laboratory and in the Department of Preventive Medicine.

Richard E. Beckwith received a Ph.D. degree in Mathematical Statistics from Purdue University in May, 1959. He has been a senior research engineer at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, following his resignation from the staff of Purdue University in September, 1958.

Allan Birnbaum has been appointed Associate Professor of Mathematical Statistics in the Department of Mathematics and the Institute of Mathematical Sciences at New York University.

Austin J. Bonis received his Ph.D. from the George Washington University in June 1959, and is now continuing his work as Deputy Director, Analysis Division, in the Office of the Assistant Secretary of Defense for Manpower and Personnel in the Department of Defense. Dr. Bonis has prepared a solution book to Dr. S. Kullback's *Information Theory and Statistics* (J. Wiley). A limited number of copies are on sale at the George Washington University Book Store at $6.25 each.

Ralph A. Bradley has joined the faculty of The Florida State University at Tallahassee, Florida. He will be Chairman of the Department of Statistics, at the University. During the past nine years, Dr. Bradley has been Professor of Statistics and consultant to the Agricultural Experiment Station at the Virginia Polytechnic Institute.

Beginning in September 1959, Dr. Irwin D. J. Bross will head the Department of Statistics at Roswell Park Memorial Institute.

Professor A. Clifford Cohen, Jr., has been named Director of the newly established University of Georgia Institute of Statistics in Athens, Georgia.

Martin Fox has accepted a position as Assistant Professor in the Department of Statistics at Michigan State University.

Irwin Guttman has been appointed Associate Professor of Mathematical Statistics, Department of Mathematics, McGill University.

Mr. Robert H. Hoskins will receive the title of Associate Group Actuary, effective September 1, 1959.

Richard C. Kao, Senior Mathematician, System Development Corporation, Santa Monica, California, is now a Senior Associate, Planning Research Corporation, Los Angeles 24, California.

Margaret P. Martin has recently joined the staff in the Department of Biostatistics of the School of Hygiene and Public Health of Johns Hopkins University as Associate Professor. She was previously with the State University of New York.

Dr. Hugh J. Miser, formerly Deputy Assistant for Operations Analysis at Hq US Air Force, has joined the staff of the Research Triangle Institute of Durham, North Carolina, as Head of its Operational Sciences Laboratory.

M. C. Pike has been appointed to the post of assistant in Mathematical Statistics.

Dr. K. C. S. Pillai has resumed his duties at the Statistics Office of the United Nations at New York after spending more than three years in the Philippines where he was United Nations Senior Statistical Adviser in Mathematical Statistics and Visiting Professor of Statistics at the Statistical Center, University of the Philippines, Manila.

William K. Robinson has taken the position of Vice-President and Actuary with the First National Life Insurance Company of Phoenix, Arizona.

The Board of Directors of the Mathematical Centre at Amsterdam has advised us of the death of David van Dantzig, Professor at the University of Amsterdam, Member of the Board and Head of the Departments of Mathematical Statistics and Applied Mathematics of the Mathematical Centre. He passed away at the age of 58 years on July 22nd 1959 after a sudden heart attack.

James K. Yarnold has recently joined the staff of General Analysis Corporation, Arizona Office.

## NEW MEMBERS

*The following persons have been elected to membership in The Institute*

May 7, 1959, to August 10, 1959

**Barr, David R.,** M.S., (Miami University); Student, U. S. Air Force Institute of Technology, Wright-Patterson AFB, Ohio, with duty station State University of Iowa, Iowa City, Iowa; *Box 608, Iowa City, Iowa.*

**Basmann, Robert L.,** Ph. D., (Iowa State College); Operations Research Analyst, *Hanford Laboratories Operation, Operations Research and Synthesis Operation, General Electric Company, Hanford Atomic Products Operation, Richland, Washington.*

**Bhat, Bellyar Ramdas,** M.A. (Karnatak University) Lecturer in Statistics, Karnatak University, Dharwar, India; *Department of Statistics, University of California, Berkeley 4, California.*

**Bobb, James C.,** B.S., (Carnegie Institute of Technology); Statistical Engineer, *Betz Laboratories, Inc., Gillingham and Worth Streets, Philadelphia 24, Pa.*

**Booker, Aaron H.,** M.A., (North Texas State College); Graduate Student, Iowa State University, Ames, Iowa; *593 Pammel Court, Ames, Iowa.*

**Bradford, Clarence H.,** M.A., (University of Chicago); Associate Director, *Army Institute Project, The University of Chicago, 5757 South Drexel Avenue, Chicago 37, Illinois.*

**Brown, Bradford S.,** M.S., (University of Illinois); Service Engineer, *Engineering Depart-*

*ment, E. I. du Pont de Nemours and Company, Buffalo Avenue and Chemical Road, Niagra Falls, New York.*

**Bryson, Marion R.**, Ph.D., (Iowa State College); Research Associate, Duke University Office of Ordnance Research, Durham, North Carolina; *Box EM, Duke Station, Durham, North Carolina.*

**Charles, Gerald T.**, B.S., (Roosevelt University); Statistician, Remington Rand Univac, 1902 West Minnehaha Avenue, St. Paul, Minnesota; *1402 North Dupont Avenue, Minneapolis 11, Minnesota.*

**Chen, Robert J.T.**, M.S., (Oklahoma State University); Graduate Student, Oklahoma State University, Stillwater, Oklahoma; *Statistical Laboratory, Oklahoma State University, Stillwater, Oklahoma.*

**Chu, Herbert H.**, M.S., (Oklahoma State University); Graduate Student, *Statistical Laboratory, Oklahoma State University, Stillwater, Oklahoma.*

**Denny, John, L., Jr.**, B.A., (Stanford University); Research Assistant, University of California, Berkeley, California; *470 Panoramic Way, Berkeley 4, California.*

**Douglas, Alan W.**, M.Sc., (McGill University); Assistant, Biometrics Unit, Department of Plant Breeding, Cornell University, 337 Warren Hall, Ithaca, New York; *137 Northview Road, Ithaca, New York.*

**Dubey, Satya D.**, B.SC., (Patna University); Special Graduate Research Assistant, *Department of Statistics, Michigan State University, East Lansing, Michigan.*

**Eicker, Friedhelm**, Dr.rer.nat., (University of Mainz); Research Associate, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

**Ericson, William A.**, M.A., (University of Pennsylvania); Student and Research Assistant, Mathematics for Application to Business, Harvard Business School; *1039 Massachusetts Avenue, Apartment 10-B, Cambridge 39, Mass.*

**Fels, Eberhard M.**, Dr.oec.publ., (University of Munich); Associate Professor of Mathematics, University of Pittsburgh, Pittsburgh 13, Pennsylvania; *6566 Forbes Avenue, Pittsburgh 17, Pa.*

**Ferruccio, George Spadaro**; Ph.D., (Polytechnic Institute of Milan); Senior Research Engineer, *Rocketdyne-Div. NAA, 6633 Canoga Ave., Canoga Park, California.*

**Finch, Robert H., Jr.**, B.A., (University of Toledo), Mathematical Statistician, Bureau of the Census, Washington 25, D. C.; *4922 Deal Drive, Washington 21, D. C.*

**Galliher, Herbert Parrish, Jr.**, Ph.D., (Yale University) Assistant Director, Operations Research Center, Mass. Institute of Technology, 77 Massachusetts Avenue, Cambridge 39, Mass.; *5 Barney Hill Road, Wayland, Massachusetts.*

**Giblin, John Martin**; M.S., (John Carroll University); Senior Programmer, *Standard Oil of Ohio, 554A Guildhall Building, Cleveland, Ohio.*

**Glick, Irving I.**, A.B., (Johns Hopkins University); Mathematician, U. S. Naval Ordnance Laboratory, Silver Spring, Maryland; *9408 Adelphi Road, Hyattsville, Maryland.*

**Greene, John A.**, M.S., (Mass. Institute of Technology); Engineer, Mathematics and Statistics Group, *E. I. du Pont de Nemours Company, Wilmington, Delaware.*

**Gross, Alan John**, M.A., (U.C.L.A.); Research Fellow, *Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina.*

**Hanson, David L.**, B.S., (Mass. Inst. of Tech.); Student, *Indiana University, Dept. of Mathematics, Bloomington, Indiana.*

**Harwayne, Frank**, B.A., (Brooklyn College); Consulting Actuary, *Frank Harwayne and Associates, 8 Stuyvesant Oval, New York 9, New York.*

**Holmes, Paul Thayer**, B.A., (State College of Washington); Teaching Assistant, Department of Mathematics, State College of Washington, Pullman, Washington; *303½ Campus, Pullman, Washington.*

**Hultquist, Robert A.**, Ph.D., (Oklahoma State University); Assistant Professor of Mathematics, *DePauw University, Greencastle, Indiana.*

**Jogdeo, Kumar;** M.Sc., (Poona University); Graduate Student and Research Assistant, *Department of Statistics, University of California, Berkeley 4, California.*

**Joseph, R. David,** A.B., (Miami University); Graduate Student and Research Mathematician, Cornell University and Cornell Aeronautical Laboratory, Ithaca, N. Y., and 4455 Genese Street, Buffalo 21, New York; *931 East State Street, Ithaca New York.*

**Kalman, Rudolf E.,** D.Engr.Sci., (Columbia University); Mathematician, *Research Institute for Advanced Study (RIAS) 7212 Bellona Ave., Baltimore 12, Maryland.*

**Kellerer, Hans G.,** Diplom-Mathematiker, (University of Munich); Assistant for Mathematical Statistics, Mathematisches Institut, Universitat Munchen, Munchen 22, Geschwister-Scholl-Platz 1; *Neugrunwald bei Munchen, Portenlangerstr. 25, Germany.*

**Kjelsberg, Marcus O.,** M.A., (University of Minnesota); Instructor in Biostatistics, *Tulane University Medical School, Department of Tropical Medicine and Public Health, 1430 Tulane Avenue, New Orleans 12, Louisiana.*

**Kleinman, David C.,** B.S., (University of Chicago); Graduate Student, Department of Statistics, University of Chicago, Chicago 37, Illinois; *5527 S. Everett Avenue, Chicago 37, Illinois.*

**Kodlin, Dankward,** M.D., (University of Heidelberg); M.P.H., (University of Pittsburgh); Assistant Professor, *Department of Biostatistics, University of Pittsburgh, Graduate School of Public Health, 130 DeSoto Street, Pittsburgh 13, Pennsylvania.*

**Lezcano, Ruben Dario,** Lt., B.S., (U.S.M.M.A.); Comandante, Canonero Paraguay, of the Paraguayan Navy, *Comando de la Armada, Hernandarias y Pte., Franco, Asuncion Paraguay.*

**Looney, Dorothea R.,** A.B., (Marymount College); Statistician, Polaroid Corporation, 730 Main Street, Cambridge 39, Mass.; *13 Magazine Street, Cambridge 39, Mass.*

**Lytle, Ernest J.,** Jr., Ph.D., (University of Florida); Statistician, Senior Technical Staff Member, *Radiation, Inc., Research Division, P. O. Box 6904, Orlando, Florida.*

**McCollum, Irving A.,** M.A., (Northwestern University); Assistant Professor, North Carolina College, Durham, North Carolina; *125 Nelson Street, Durham, N. C.*

**MacDonald, Eleanor J.,** A.B., (Radcliffe College); Epidemiologist Professor of Biostatistics, *University of Texas Postgraduate School of Medicine, M.D. Anderson Hospital and Tumor Institute, Branch of the University of Texas, 6723 Bertner Drive, Houston, Texas.*

**Mason, Frank D.,** M.S., (Washington State College); Student, N. C. State College, Department of Experimental Statistics, Raleigh, North Carolina.

**Mayne, Alan J.,** B.Sc., (Oxford University); Independent Research Scientist and Consultant, *82, Divinity Road, Oxford, England.*

**Michalski, John P.,** M.S., (Northwestern University), Student, Alabama Polytechnic Institute, Auburn, Alabama; *1509 West 10th Court, Panama City, Florida.*

**Miller, E. Jacques,** B.S., (California Institute of Technology); Comptroller's Staff, *C. F. Braun and Company, 1000 South Fremont Avenue, Alhambra, California.*

**Miller, Robert G.,** M.S., (New York University); Research Associate, Travelers Insurance Company, 700 Main Street, Hartford, Conn.; *86 Hilldale Road, West Hartford, Conn.*

**Murray, John E.,** A.B., (Shimer College); Student, University of Chicago, Chicago 37, Illinois; *1127 North Francis Street, South Bend 17, Indiana.*

**Nathan, Ravindra,** M.Sc., (University of Travancore); Acting Instructor, Part-time, *Mathematics Department, University of Idaho, Moscow, Idaho.*

**Nelson, William G., IV,** M.B.A., (Wharton Graduate School, University of Penn.); Accountant, E. I. du Pont de Nemours and Company, Wilmington 3, Delaware; *1109 Faun Road, Wilmington 3, Delaware.*

**Padro, Jose R.,** M.A., (The Ohio State University); Assistant Professor, University of Puerto Rico, Rio Piedras, Puerto Rico; *Department of Mathematics, St. Louis University, 221 North Grand Blvd., % Dr. Waldo A. Vezeau, St. Louis 3, Missouri.*

**Nieto de Pascual, Jose,** MS., (Iowa State College); Instructor, *Department of Statistics, Iowa State College, Ames, Iowa.*

**Perlstein, Helen W.,** B.S., (College of the City of New York); Student, Columbia Uni-

versity, Morningside Heights, New York 27, New York; *400 West 118 Street, New York 27, New York.*

**Rao, Malempati Madhusudana,** Ph.D., (University of Minnesota); Research Fellow, University of Minnesota, Vincent Hall, Minneapolis 14, Minn.

**Sawyer, Jack,** Ph.D., (Purdue University); Postdoctoral Fellow, *Department of Statistics, University of Chicago, Chicago 37, Illinois.*

**Schupp, Carl F.,** B.A., (University of Delaware); Statistician, *E. I. du Pont de Nemours and Co., Inc., Engineering Department (Louviers Building), Wilmington, Del.*

**Shugg, W. Tillar,** (Union College); Manager Marketing Services, Continental-Diamond Fibre Corp. (a division of The Budd Co.), Newark, Delaware; *Egypt Rd., RD. 2, Newark, Delaware.*

**Slotsky, Gordon J.,** B.S., (Iowa State College); Assistant Engineer, *Reliability Operations, Countermeasures, Sperry Gyroscope Company, Division of Sperry Rand Corporation, Great Neck, Long Island, New York.*

**Stevens, Roger T.,** B.A., (Union College); Senior Electronic Engineer, Electronics Systems, Inc., 105 Chauncy Street, Boston 11, Mass; *33 Hayden Lane, Bedford, Mass.*

**Sylvan, Marshall,** B.S., (Roosevelt University); Student, University of Chicago, Chicago 37, Illinois; *8921 South Harper Avenue, Chicago 19, Illinois.*

**Takács, Lajos F.,** Doctor of Mathematical Sciences (Hungarian Academy of Sciences); Assistant Professor, *Columbia University, New York 27, N. Y.*

**Thomas, Ronald Emerson,** M.A., (Queen's University); Student, *University of North Carolina, Department of Statistics, Chapel Hill, North Carolina.*

**Truax, Harry Mack,** M.S., (University of Delaware); Manager of Quality Control Section, *Production Department, Atlas Powder Company, Wilmington, Delaware.*

**Van Ryzin, John R.,** B.S., (Marquette University); Graduate Assistant, Marquette University; *1810 N. 73rd Street, Wauwatosa, Wisconsin.*

**Wang, Yin Huai,** M.S., (University of Tennessee); Half-time Teaching Assistant, Department of Mathematics and Statistics at Purdue University, Lafayette, Indiana; *103 W. Stadium Avenue, West Lafayette, Indiana.*

**Wasan, Madanlal T.,** M.A., (University of Bombay); Assistant in Department of Mathematics, Department of Mathematics, University of Illinois, Urbana, Illinois; *274 Altgeld Hall, University of Illinois, Urbana, Illinois.*

**Wells, William J.,** M.S., (Univ. of Michigan); Mathematician, *University of Chicago, Institute for Air Weapons Research, Museum of Science and Industry, Chicago 37, Illinois.*

**Wells, William T.,** Ph.D., (North Carolina State College); Statistician, Research Triangle Institute, 505 West Chapel Hill Street, Durham, North Carolina; *707-D Daniels Street, Raleigh, North Carolina.*

**Whinston, Andrew Bernard,** B.A., (University of Michigan); Student, *Graduate School of Industrial Administration, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.*

**Wirjosudirdjo, Sunardi,** M.S., (University of Indonesia); Student, Department of Mathematics, University of Illinois, Urbana, Illinois; *1110 West Main Street, Urbana, Illinois.*

**Wood, Ross J.,** M.B.A., (University of Michigan); Statistician, Remington Rand Univac, St. Paul, Minnesota; *1340 E. County Line Rd., S., St. Paul 6, Minnesota.*

**Wright, Benjamin D.,** Ph.D., (University of Chicago); Assistant Professor of Educational Psychology, *University of Chicago, Graduate School of Education, 5835 Kimbark Avenue, Chicago 37, Illinois.*

**Yao, Ying,** B.A., (Barnard College); Graduate Student, University of Chicago, Chicago 37, Illinois; *138-09 76th Avenue, Flushing 67, New York.*

**Young, Shirley,** B.A., (Pennsylvania State University); Mathematician, National Bureau of Standards, Connecticut and Van Ness, Washington 25, D. C.; *3706 Harrison, N. W., Washington 25, D. C.*

**Zubin, Joseph,** Ph.D., (Columbia University); Principal Research Scientist, (Biometrics), and Prof. of Psychol. (Columbia); *New York State Department of Mental Hygiene, 722 West 168 Street, New York 32, New York.*

## MEETING OF ISI

The 32nd Session of the International Statistical Institute will be held in Tokyo from May 30 to June 9, 1960. The program covers a wide variety of topics in theoretical and applied statistics. Detailed information about registration, program, etc. may be obtained from either *Mr. Masao Goto, Organizing Committee of the 32nd I.S.I. Session, ℅ Statistical Standards Bureau Administrative Management Agency, 5 Sannen-cho, Chiyoda-ku, Tokyo, Japan*, or *Permanent Office, International Statistical Institute, 2 Oostduinlaan, The Hague, Netherlands.*

## NATIONAL BUREAU OF STANDARDS POSTDOCTORAL RESEARCH ASSOCIATESHIPS

Research associateships, supported by the National Bureau of Standards and awarded on recommendations of the National Academy of Sciences—National Research Council, are offered to provide young investigators of unusual ability and promise the opportunity for basic research in various branches of the physical and mathematical sciences. It is expected that approximately 20 awards may be made in a total of twenty-nine fields, of which the following are of particular interest to members of the institute: Pure and Applied Mathematics, Applied Mathematical Statistics, Numerical Analysis. These research associateships are open only to citizens of the United States, and in the foregoing fields are tenable only at the National Bureau of Standards in Washington, D. C. Applicants must have received (or completed the requirements for) a Ph.D. or Sc.D. degree, or the equivalent, in one of the fields listed above at the time of entering upon the research associateship.

The annual gross stipend will be $7510 and will be subject to income tax. Travel and moving expenses of the Research Associate and his family from place of residence to Washington, D. C. will be paid by the National Bureau of Standards. Awards will be made about April 1, 1960. Unless otherwise arranged the tenure of a research associateship may begin after July 1, 1960 and continue for one year, with provision for a vacation period.

Requests for application forms or for additional information should be addressed to the Fellowship Office, National Academy of Sciences—National Research Council, 2102 Constitution Avenue, N. W., Washington 25, D. C. *Applications for the academic year 1960–1961 must be received in the Fellowship Office no later than February 1, 1960.*

## FELLOWSHIP AND RESEARCH OPPORTUNITIES

The Division of Mathematics, National Academy of Sciences—National Research Council, has published a leaflet listing some private and government

agencies that offer support of mathematical study and research at the graduate and post-graduate levels. Copies of the leaflet are available upon application to *Division of Mathematics, National Academy of Sciences—National Research Council, 2101 Constitution Avenue, Washington 25, D. C.*

## NEW EDITORIAL STAFF FOR MTAC

The Division of Mathematics of the National Academy of Sciences—National Research Council announces that Harry Polachek, Technical Director of the Applied Mathematics Laboratory of the David Taylor Model Basin, has been appointed Chairman of the Editorial Committee for the quarterly journal *Mathematical Tables and Other Aids to Computation* effective January 1959. He succeeds C. B. Tompkins of the University of California at Los Angeles, who held the post since November 1954. The other members of the Editorial Committee are: C. C. Craig, A. Fletcher, E. Isaacson, D. Shanks, C. V. L. Smith, A. H. Taub, C. B. Tompkins and J. W. Wrench, Jr.

Articles for publication in *Mathematical Tables and Other Aids to Computation* should be addressed to *Harry Polachek, Editor, Mathematical Tables and Other Aids to Computation, David Taylor Model Basin, Washington 7, D. C.*

Information on subscriptions may be obtained from *National Academy of Sciences, Printing and Publishing Office, 2101 Constitution Avenue, N. W., Washington 25, D. C.*

## FLORIDA STATE UNIVERSITY DEPARTMENT OF STATISTICS

The Florida State University at Tallahassee, Florida has established a Department of Statistics effective July, 1959. The initial faculty will consist of Ralph A. Bradley, Chairman, John L. Bagg and Lonnie L. Lasman. Programs of study leading to the Bachelor of Science and Master of Science degrees in statistics will be initiated in the Fall 1959 semester and advanced graduate work will be developed in the near future. The Department of Statistics will provide university-wide training and consulting. Inquiries regarding the program will be welcomed and some assistantships will be available for graduate students.

## KANSAS STATE UNIVERSITY DEPARTMENT OF STATISTICS

July 1, 1959, Kansas State University at Manhattan, Kansas, established a new Department of Statistics. A Statistical Laboratory was organized at Kansas State in 1946 and will continue as a consulting and computing unit sponsored by the Agricultural Experiment Station.

The Department of Statistics presently has a five-man staff: Robert S. Cochran, Arlin M. Feyerherm, H. C. Fryer, Gary F. Krause, and Stanley Wearden, plus graduate and research assistants. A Master's degree in statistics has been offered for a number of years through the Department of Mathematics. It is planned that a Ph.D. in statistics will be offered soon.

## RUTGERS STATISTICS CENTER

Rutgers University has established a Statistics Center which will be the State University's central unit for research and teaching in the field of statistics. It will continue to offer at Rutgers the Master's Degree program and in addition will have responsibility for a program of study leading to the Ph.D. in Applied and Mathematical Statistics.

The Center, which is under the administration of Dr. Marion A. Johnson, dean of the Graduate School, will be directed by Dr. Ellis R. Ott, who has previously served as professor of Mathematics and Chairman of the Mathematics Department of University College, as well as Chairman of the Rutgers program in Applied and Mathematical Statistics. Director of Research at the Center will be Dr. Martin B. Wilk. The staff will also include Dr. Roger S. Pinkham, Dr. Mason E. Wescott and Harold F. Dodge.

The Statistics Center will be responsible for programs of research in statistical theory and methodology and in ways of promoting the more effective use of statistics in science, industry, education and other areas.

The Center will also make available, both within and without the University, consultative services with respect to the use of statistics and statistical techniques in research or other experiments and surveys, and with respect to the analysis and presentation of data.

Since 1952, Rutgers has had course work leading to a Master's degree in the field of Applied and Mathematical Statistics. Through June, 1959, a total of 85 Master's degrees in applied and mathematical statistics had been awarded, mostly to full-time employees in nearby industries. The present development constitutes a major extension of the Rutgers Statistics Program plus an administrative reorganization.

The establishment of the Statistics Center is predicated on the belief that a group within the University charged with instructional, research, and consultative functions in the field will both advance and strengthen the University's program in statistics, and also further the effective use of statistical tools and techniques by the many members of the University staff who, though not specialists in statistics, must of necessity use statistics in their own work.

## PUBLICATIONS RECEIVED

*Annuario Estadistico de España, Edicion Manual*, Presidencia del Gobierno, Instituto Nacional de Estadistica, Ferraz 41, Madrid, Spain, 1959, 977 pp.

Bush, Robert R., and Estes, William K., Editors, *Studies in Mathematical Learning Theory*, Stanford University Press, Stanford, California, 1959, 432 pp., $11.50

*Cahiers du Séminaire D'Économétrie, No. 5, Production, investissements et productivité*, Éditions du Centre National de la Recherche Scientifique, 13 Quai Anatole-France, Paris-7ᵉ, France, 1959, 178 pp.

*Statistical Handbook, 1958*, Central Statistical Office, Peoples Republic of Bulgaria, Sofia, 1959, 223 pp.

# TECHNOMETRICS

*A Journal of Statistics*

*for the Physical, Chemical and Engineering Sciences*

Vol. 1, No. 3, Contents, August 1959



J. Stuart Hunter, *Editor*
Mathematics Research Center, U.S. Army
The University of Wisconsin
Madison, Wisconsin